Beam Training and Tracking for Extremely Large-Scale MIMO mmWave/THz Communications

Journal:	*IEEE Transactions on Wireless Communications*
Manuscript ID	Paper-TW-Mar-23-0452
Manuscript Type:	Original Transactions Paper
Date Submitted by the Author:	28-Mar-2023
Complete List of Authors:	Chen, Kangjian; Southeast University, School of Information Science and Engineering Qi, Chenhao; Southeast University, School of Information Science and Engineering Wang, Cheng-Xiang; Southeast University, National Mobile Communications Research Laboratory; Purple Mountain Laboratories, Pervasive Communication Research Center Li Ye, Geoffrey; Imperial College London, Electrical and Electronic Engineering
Keyword:	

Beam Training and Tracking for Extremely Large-Scale MIMO mmWave/THz 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 millimeter wave (mmWave) and terahertz (THz) communication systems, where extremely large-scale multiple-input-multiple-output (MIMO) with partially-connected hybrid combining structures is adopted. First, we propose a two-stage hybrid-field beam training scheme for both the near field and 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

Millimeter wave (MmWave) and terahertz (THz) have attracted great interests due to their abundant spectrum resources. In the existing fifth generation wireless communications, massive multiple-input-multiple-output (MIMO) has been integrated into the mmWave communications to improve the spectral efficiency by exploiting the spatial degree of freedom [2]–[4]. For future sixth generation wireless communications, extremely large-scale MIMO (XL-MIMO) with far more antennas than the existing massive MIMO is considered [5]–[9].

Due to large power consumption, the fully-digital structure that allocates each antenna with a dedicated radio frequency (RF) chain is impractical for large antenna arrays [10]–[12].

Consequently, the hybrid structure, where a small number of RF chains are connected to a large number of antennas, is developed for XL-MIMO systems [13]. According to the ways how the RF chains are connected to the antennas, the hybrid structure can generally be divided into two categories, including fully-connected structure and partially-connected structure. Although the fully-connected structure can achieve better spectral efficiency than the partially-connected 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 [14]–[16].

One important difference between XL-MIMO and the existing massive MIMO is the channel features. Depending on the distance between the user and the BS, the radiation field can be divided into the near field and the far field, bounded by the Rayleigh distance [17]–[19]. 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 systems, especially when a user is close to the BS. In this context, channel state information (CSI) acquisition methods for XL-MIMO should take both the near-field and far-field effects into consideration.

In general, CSI acquisition includes channel estimation and beam training [20]. Channel estimation usually focuses on efficient estimation of the high-dimension channel matrix by exploiting advanced signal processing techniques, such as compressed sensing, while the beam training can avoid the estimation of the high-dimension channel matrix and obtain considerable beamforming gain, especially in low signal-to-noise ratio (SNR). In [13], 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. In [21], hierarchical codebooks for both the far field and the near field are designed based on the uniformly quantized space. In [22], a chirp-based hierarchical codebook is designed based on the expressions of the near-field channels in the slope-intercept domain.

One common challenge on the 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 for angle estimation of far-field channels [23]–[25]. In [23], beam refinement with a closed-form solution based on the monopulse signals is developed. In [24], an efficient angle-of-arrival estimator is

designed by approximating the power of the array response as a Gaussian function. In [25], an auxiliary beam pair is designed to provide high-resolution estimate for the angle-of-departure of the channel. These 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 heavy burden of training overheads if still using existing methods. An efficient way to reduce the training overhead is the beam tracking, which exploits the correlations of the channels at different time slots to narrow down the sets of training candidates. There are a variety of beam tracking schemes for the far-field channels [26]–[29]. In [26], well-designed pairs of auxiliary beams can capture the angle variations. The beam zooming-based beam tracking scheme in [27] exploits the delay-phase precoding structure to flexibly control the angular coverage of frequency-dependent beams. In the adaptive tracking framework in [28], beam direction is updated according to measurements of the current data beam. The grid-based hybrid tracking scheme in [29] searches the surroundings of the former beam and selects the best beam according to the changing trend of the previously used beams.

To the best knowledge of the authors, so far there has been no literature working on effective beam training, refinement, and tracking for XL-MIMO systems with near-field effect. To fill in the gaps, we investigate these issues for XL-MIMO systems with partially-connected structures in this paper. 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 a subarray 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, channel parameters can be estimated with high resolution.

- 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, a 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.
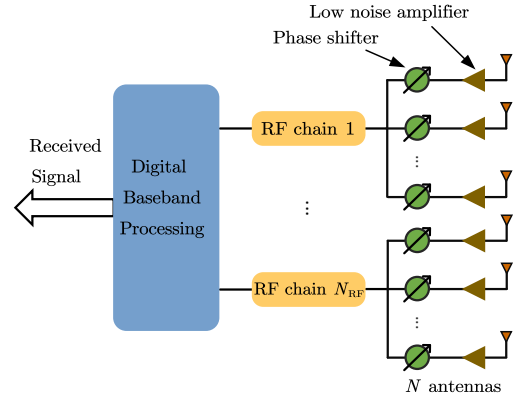

Fig. 1.   Illustration of a partially-connected hybrid combining structure.

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. $(\cdot)^{\mathrm{T}}$ and $(\cdot)^{\mathrm{H}}$ denote the transpose and conjugate transpose (Hermitian), respectively. $[\boldsymbol{A}]_{:,m}$ denotes the mth column of a matrix $\boldsymbol{A}$. 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 $\mathrm{blkdiag}\{\cdot\}$ denote the floor operation and the block diagonalization operation, respectively. Moreover, $\mathcal{O}$ and $\mathrm{mod}(\cdot)$ 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 hybrid structure has far more antennas than RF chains, i.e., $N \gg N_{\mathrm{RF}}$. The ULA is formed by N_{RF} non-overlapping subarrays, where each subarray has $M = N/N_{\mathrm{RF}}$ 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 analog combining and digital combining at the BS side while a single-antenna user is considered for simplification.

During uplink beam training, the pth transmit signal by the user is denoted as x_p for $p = 1, 2, \ldots, P$, where P is the signal length. The channel vector between the user and the BS is denoted as $\boldsymbol{h} \in \mathbb{C}^N$. Then the signal after the analog

combining at the BS side can be expressed as [15]

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

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

Generally, two kinds of channels, including 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 [17]–[19]. The commonly used boundary distance to distinguish the near field and the far field is the Rayleigh distance

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

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-filed channel; otherwise, it is the near-filed 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 $Z = N\lambda^2/2$.

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 nth antenna is $(0, \delta_n \lambda)$, where $\delta_n \triangleq (2n - N - 1)/4$ for $n = 1, 2, \ldots, N$. The coordinate of the user is denoted as $\boldsymbol{p}_1 = (r_1 \cos\theta_1, r_1 \sin\theta_1)$, 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 lth path for $l \geq 2$ is denoted as $\boldsymbol{p}_l = (r_l \cos\theta_l, r_l \sin\theta_l)$, where r_l is the distance between the scatterer on the lth path and the coordinate origin, and $\theta_l \in [-\pi/2, \pi/2]$ is the angle of the scatterer on the lth path relative to the positive x-axis. The distance between $\boldsymbol{p}_l$ and the nth antenna can be expressed as

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

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

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

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

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

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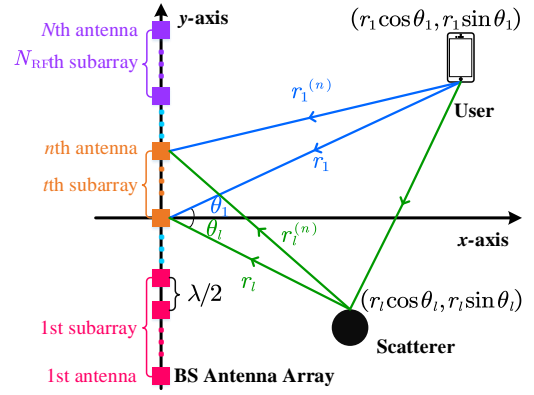

Fig. 2. Illustration of a near-field channel model.

If $r_l > Z$, $r_l^{(n)}$ in (3) can be simplified as $r_l^{(n)} \approx r_l - \Omega_l \delta_n \lambda$ because $\delta_n \lambda / r_l \approx 0$ and $\sqrt{1 + \epsilon} \approx 1 + \frac{1}{2}\epsilon$ [13]. As a result, $\boldsymbol{\alpha}(N, \Omega_l, r_l)$ in (5) can be approximated as

$$\boldsymbol{\alpha}(N, \Omega_l, r_l) \approx \frac{1}{\sqrt{N}}\left[e^{j2\pi\Omega_l\delta_1}, \ldots, e^{j2\pi\Omega_l\delta_N}\right]^{\mathrm{T}}. \qquad (6)$$

Removing a constant factor $e^{j2\pi\Omega_l\delta_1}$ from all the entries of $\boldsymbol{\alpha}(N, \Omega_l, r_l)$ in (6), we have

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

In fact, $\boldsymbol{\beta}(N, \Omega_l)$ 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 simplify $r_l^{(n)}$ 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 $r_l^{(n)} \geq 0.5\sqrt{D^3/\lambda}$ [13], [18]. Note that the ratio of Z to $0.5\sqrt{D^3/\lambda}$ is $4\sqrt{N}$, which indicates that the latter is much smaller than the former, especially for XL-MIMO systems. For example, if $N = 1024$, $0.5\sqrt{D^3/\lambda}$ is only $1/128$ of Z, which is almost negligible. Therefore, in this work, we focus on the near field with $r_l^{(n)} \geq 0.5\sqrt{D^3/\lambda}$. Then $\boldsymbol{\alpha}(N, \Omega_l, r_l)$ in (5) can be approximated as

$$\boldsymbol{\alpha}(N, \Omega_l, r_l) \approx \frac{1}{\sqrt{N}}\left[e^{j2\pi(\Omega_l\delta_1 - \rho_l\delta_1^2)}, \ldots, e^{j2\pi(\Omega_l\delta_N - \rho_l\delta_N^2)}\right]^{\mathrm{T}} \qquad (8)$$

where $\rho_l \triangleq \frac{\lambda(1 - \Omega_l^2)}{2r_l}$. Define $b_l \triangleq \Omega_l + \rho_l(N + 1)/2$ and $k_l \triangleq -\rho_l/2$. Removing a constant factor $e^{j2\pi(\Omega_l(\delta_1 - 1/2) - \rho_l(N+1)^2/16)}$ from all the entries of $\boldsymbol{\alpha}(N, \Omega_l, r_l)$ in (6), we have

$$e^{-j2\pi(\Omega_l(\delta_1 - 1/2) - \rho_l(N+1)^2/16)}\boldsymbol{\alpha}(N, \Omega_l, r_l)$$
$$\approx \frac{1}{\sqrt{N}}\left[e^{j\pi(k_l + b_l)}, \ldots, e^{j\pi(k_l n^2 + b_l n)}, \ldots, e^{j\pi(k_l N^2 + b_l N)}\right]^{\mathrm{T}}$$
$$\triangleq \boldsymbol{\gamma}(N, \Omega_l, r_l). \qquad (9)$$

Note that $\boldsymbol{\beta}(N, \Omega_l)$ is a special case of $\boldsymbol{\gamma}(N, \Omega_l, r_l)$ with $\rho_l = 0$. Therefore, $\boldsymbol{\gamma}(N, \Omega_l, r_l)$ 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.

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 $\boldsymbol{C}_{\mathrm{f}} \in \mathbb{C}^{N \times Q}$ based on the far-field channel steering vectors [3], where Q denotes the quantized number of the angle and the qth column of $\boldsymbol{C}_{\mathrm{f}}$ is denoted as

$$[\boldsymbol{C}_{\mathrm{f}}]_{:,q} \triangleq \boldsymbol{\beta}\left(N, \frac{2q-1-Q}{Q}\right) \qquad (10)$$

for $q = 1, 2, \ldots, Q$. Then we establish a near-field codebook $\boldsymbol{C}_{\mathrm{n}} \in \mathbb{C}^{N \times (QS)}$. 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 qth angle sample is $\Theta_q = (2q-1-Q)/Q$ for $q = 1, 2, \ldots, Q$. The sth distance sample is

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

for $s = 1, 2, \ldots, S$, where β is a factor to adjust the coherence of the adjacent codewords [13]. By setting $d_{(Q+1)/2,S} = 0.5\sqrt{D^3/\lambda}$, we have $\beta = \sqrt{\frac{N}{2S^2}}$. Then (11) is rewritten as

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

Then we let $\boldsymbol{C}_{\mathrm{n}} \triangleq \{\boldsymbol{C}_1, \boldsymbol{C}_2, \ldots, \boldsymbol{C}_Q\}$, where the sth column of $[\boldsymbol{C}_q]$ is $[\boldsymbol{C}_q]_{:,s} = \boldsymbol{\alpha}(N, \Theta_q, d_{q,s})$. Accordingly, the hybrid-field codebook is defined as

$$\boldsymbol{C}_{\mathrm{h}} \triangleq \{\boldsymbol{C}_{\mathrm{n}}, \boldsymbol{C}_{\mathrm{f}}\} \in \mathbb{C}^{N \times (QS+Q)}. \qquad (13)$$

Based on (1), we have

$$y_p = [\boldsymbol{C}_{\mathrm{h}}]_{:,p}^{\mathrm{H}} \boldsymbol{h} x_p + [\boldsymbol{C}_{\mathrm{h}}]_{:,p}^{\mathrm{H}} \boldsymbol{\eta} \qquad (14)$$

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

$$\widetilde{p} = \arg \max_{p=1,2,\ldots,QS+Q} \left| [\boldsymbol{C}_{\mathrm{h}}]_{:,p}^{\mathrm{H}} \boldsymbol{h} \right|. \qquad (15)$$

To solve (15), we need to test all the codewords in $\boldsymbol{C}_{\mathrm{h}}$ one by one, which is called the hybrid-field beam sweeping. To perform the hybrid-field beam sweeping, $(QS + Q)$ 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.

We define an analog combiner as

$$\boldsymbol{W} \triangleq \mathrm{blkdiag}\{\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_{N_{\mathrm{RF}}}\} \qquad (16)$$

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

$$\max_{\boldsymbol{W}, \boldsymbol{v}} \quad \left| \boldsymbol{v}\boldsymbol{W}\boldsymbol{h} \right|$$
$$\mathrm{s.t.} \quad \|\boldsymbol{v}\boldsymbol{W}\|_2 = 1, \left| [\boldsymbol{w}_t]_m \right| = 1 \qquad (17)$$

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

$$\max_{\boldsymbol{W}, \boldsymbol{v}} \left| \boldsymbol{v}\boldsymbol{W}\boldsymbol{\alpha}(N, \Omega, r) \right| \qquad (18a)$$
$$\mathrm{s.t.} \quad \|\boldsymbol{v}\boldsymbol{W}\|_2 = 1, \ \left| [\boldsymbol{w}_t]_m \right| = 1 \qquad (18b)$$

for $m = 1, 2, \ldots, M$ and $t = 1, 2, \ldots, N_{\mathrm{RF}}$. In (18a), the subscript "l" is omitted for the simplification for the rest of this paper.

According to the Cauchy-Schwartz inequality, we have

$$\left| \boldsymbol{v}\boldsymbol{W}\boldsymbol{\alpha}(N, \Omega, r) \right| \leq \|\boldsymbol{v}\|_2 \|\boldsymbol{W}\boldsymbol{\alpha}(N, \Omega, r)\|_2. \qquad (19)$$

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

The design of $\boldsymbol{W}$ can be formulated as

$$\max_{\boldsymbol{W}} \left\| \boldsymbol{W}\boldsymbol{\alpha}(N, \Omega, r) \right\|_2$$
$$\mathrm{s.t.} \ \left| [\boldsymbol{w}_t]_m \right| = 1 \qquad (20)$$

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

$$\max_{\boldsymbol{w}_t} \left| \boldsymbol{w}_t \boldsymbol{G}_t \boldsymbol{\alpha}(N, \Omega, r) \right|$$
$$\mathrm{s.t.} \ \left| [\boldsymbol{w}_t]_m \right| = 1 \qquad (21)$$

for $m = 1, 2, \ldots, M$, where we define $\boldsymbol{G}_t \triangleq [\boldsymbol{0}_{M \times (t-1)M}, \boldsymbol{I}_M, \boldsymbol{0}_{M \times (N_{\mathrm{RF}}-t)M}]$. The optimal solution of (21) is

$$\overline{\boldsymbol{w}}_t = \sqrt{N}\big(\boldsymbol{G}_t \boldsymbol{\alpha}(N, \Omega, r)\big)^{\mathrm{H}} \qquad (22)$$

for $t = 1, 2, \ldots, N_{\mathrm{RF}}$.

Note that $\boldsymbol{G}_t \boldsymbol{\alpha}(N, \Omega, r)$ is essentially the channel steering vector between the user and the tth subarray. From (2), the whole 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 $1/N_{\mathrm{RF}}^2$ of the Rayleigh distance of the whole array at the BS, which implies that a user in the near field of the whole array might be in the far field

of a subarray. For example, if $N = 256$, $\lambda = 0.003$ m, and $N_{\mathrm{RF}} = 4$, the Rayleigh distance of the whole 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 whole array. Based on the above discussion, we aim 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

$$\Gamma_{\max} = \max\{1 - N_{\mathrm{RF}}/\sqrt[4]{2N}, 0\}. \tag{23}$$

Lemma 2: Denote the sine result of the angle that points from the center of the tth subarray to the user as Ψ_t and denote the beam center of $\boldsymbol{G}_t \boldsymbol{\alpha}(N, \Omega, r)$ 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 $N_{\mathrm{RF}} = 4$, the maximum beam gain loss is only 16%. From **Lemma 2**, considerable beamforming gain can be obtained if the tth subarray receives the signal from channel $\boldsymbol{G}_t \boldsymbol{\alpha}(N, \Omega, r)$ with $\boldsymbol{\beta}(M, \Psi_t)$.

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

$$\widehat{\boldsymbol{w}}_t = \sqrt{M}\boldsymbol{\beta}(M, \Psi_t)^{\mathrm{H}} \tag{24}$$

for $t = 1, 2, \ldots, N_{\mathrm{RF}}$ to approximate the near-field channel steering vectors in (22). Therefore, we can use (24) for both the far-field and near-field channels. By substituting (24) into (16), we express the designed analog combiner for (18) as

$$\widehat{\boldsymbol{W}} = \mathrm{blkdiag}\{\widehat{\boldsymbol{w}}_1, \widehat{\boldsymbol{w}}_2, \ldots, \widehat{\boldsymbol{w}}_{N_{\mathrm{RF}}}\}. \tag{25}$$

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

$$\max_{\boldsymbol{v}} \boldsymbol{v}\widehat{\boldsymbol{W}}\boldsymbol{\alpha}(N, \Omega, r) \tag{26a}$$

$$\mathrm{s.t.} \ \|\boldsymbol{v}\|_2 = 1/\sqrt{N}. \tag{26b}$$

Note that (18a) can be rewritten as (26a) because we can always adjust the phase of $\boldsymbol{v}$ so that $\boldsymbol{v}\widehat{\boldsymbol{W}}\boldsymbol{\alpha}(N, \Omega, r)$ is real positive and the maximum of $|\boldsymbol{v}\widehat{\boldsymbol{W}}\boldsymbol{\alpha}(N, \Omega, r)|$ is still achieved. The optimal $\boldsymbol{v}$ for (26) is

$$\widehat{\boldsymbol{v}} = \frac{\left(\widehat{\boldsymbol{W}}\boldsymbol{\alpha}(N, \Omega, r)\right)^{\mathrm{H}}}{\sqrt{N}\|\widehat{\boldsymbol{W}}\boldsymbol{\alpha}(N, \Omega, r)\|_2}. \tag{27}$$

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

$$\boldsymbol{\Phi} = \{\sqrt{M}\boldsymbol{\beta}(M, \Phi_1), \sqrt{M}\boldsymbol{\beta}(M, \Phi_2), \ldots, \sqrt{M}\boldsymbol{\beta}(M, \Phi_M)\}, \tag{28}$$

where

$$\Phi_m = (2m - 1 - M)/M \tag{29}$$

for $m = 1, 2, \ldots, M$. In fact, the DFT codebook equally samples the full space $[-1, 1]$ by M angles, where the mth angle is Φ_m. For the tth subarray, the index of the codeword in $\boldsymbol{\Phi}$ best fitting for the channel is

$$\widetilde{m}_t = \arg \min_{m=1,2,\ldots,M} |\Phi_m - \Psi_t|. \tag{30}$$

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

$$\widetilde{\boldsymbol{W}} = \mathrm{blkdiag}\{\widetilde{\boldsymbol{w}}_1, \widetilde{\boldsymbol{w}}_2, \ldots, \widetilde{\boldsymbol{w}}_{N_{\mathrm{RF}}}\}, \tag{31}$$

$$\widetilde{\boldsymbol{v}} = \frac{\left(\widetilde{\boldsymbol{W}}\boldsymbol{\alpha}(N, \Omega, r)\right)^{\mathrm{H}}}{\sqrt{N}\|\widetilde{\boldsymbol{W}}\boldsymbol{\alpha}(N, \Omega, r)\|_2}, \tag{32}$$

where

$$\widetilde{\boldsymbol{w}}_t = [\boldsymbol{\Phi}]_{:,\widetilde{m}_t}^{\mathrm{H}} \tag{33}$$

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

Note that for the partially-connected structure, each subarray is solely connected to an RF chain, which indicates that each RF chain can support an independent beam training based on a subarray. As a result, we need totally M times of beam training. In particular, 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.

Now we propose a two-stage hybrid-field beam training scheme. In the first stage, each subarray independently uses M far-field channel steering vectors for analog combining, where the mth combiner according to (28) is

$$\overline{\boldsymbol{W}}_m = \mathrm{blkdiag}\{\sqrt{M}\boldsymbol{\beta}(M, \Phi_m)^{\mathrm{H}}, \ldots, \sqrt{M}\boldsymbol{\beta}(M, \Phi_m)^{\mathrm{H}}\}. \tag{34}$$

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

$$\boldsymbol{z}_m = \overline{\boldsymbol{W}}_m \boldsymbol{h} x_m + \overline{\boldsymbol{W}}_m \boldsymbol{\eta} \tag{35}$$

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

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

$$[\boldsymbol{C}_{\mathrm{h}}]_{:,p} = \begin{cases} \boldsymbol{\alpha}\left(N, \Theta_{\bar{q}}, d_{\bar{q},\bar{s}}\right), & p \leq QS, \\ \boldsymbol{\beta}\left(N, \frac{2(p-QS)-1-Q}{Q}\right), & p > QS, \end{cases} \tag{36}$$

where

$$\bar{q} = \lceil p/S \rceil \ \text{and} \ \bar{s} = p - (\bar{q} - 1)S. \tag{37}$$

Suppose $[\boldsymbol{C}_{\mathrm{h}}]_{:,p}$ is the codeword best fitting for the channel. We will design analog combiner $\boldsymbol{F}_p$ and digital combiner $\widetilde{\boldsymbol{v}}_p$ so that the phase differences of subarrays induced by the near-field effect can be compensated and the beam gain of $\widetilde{\boldsymbol{v}}_p\boldsymbol{F}_p$ can approach that of $[\boldsymbol{C}_{\mathrm{h}}]_{:,p}$. Replacing $\boldsymbol{\alpha}(N, \Omega, r)$ in (18) by $[\boldsymbol{C}_{\mathrm{h}}]_{:,p}$, we can obtain $\widetilde{\boldsymbol{w}}_t^{(p)}$ via (33). Similar to (31) and

---

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

1: **Input:** N, N_{RF}, M, S, λ.
2: **First Stage:**
3: Obtain $\boldsymbol{z}_m$, $m = 1, 2, \ldots, M$ via (35).
4: **Second Stage:**
5: **for** $p = 1, 2, \ldots, QS + Q$ **do**
6:     Obtain $[\boldsymbol{C}_{\text{h}}]_{:,p}$ via (36).
7:     Obtain $\boldsymbol{F}_p$ and $\widetilde{\boldsymbol{v}}_p$ via (38) and (39), respectively.
8:     Obtain $\widetilde{\boldsymbol{z}}_p$ via (42).
9:     Obtain $\widetilde{y}_p$ via (41).
10: **end for**
11: Obtain $\widetilde{p}$ via (43).
12: Obtain $\widetilde{\Omega}$ and $\widetilde{r}$ via (45).
13: **Output:** $[\boldsymbol{C}_{\text{h}}]_{:,\widetilde{p}}$, $\widetilde{\Omega}$ and $\widetilde{r}$.

---

(32), we design the analog combiner and the digital combiner, respectively, as

$$\boldsymbol{F}_p = \text{blkdiag}\{\widetilde{\boldsymbol{w}}_1^{(p)}, \widetilde{\boldsymbol{w}}_2^{(p)}, \ldots, \widetilde{\boldsymbol{w}}_{N_{\text{RF}}}^{(p)}\}, \quad (38)$$

$$\widetilde{\boldsymbol{v}}_p = \frac{(\boldsymbol{F}_p[\boldsymbol{C}_{\text{h}}]_{:,p})^{\text{H}}}{\sqrt{N}\|\boldsymbol{F}_p[\boldsymbol{C}_{\text{h}}]_{:,p}\|_2}. \quad (39)$$

Note that both $\boldsymbol{F}_p$ and $\widetilde{\boldsymbol{v}}_p$ can be computed offline before the beam training, which can substantially reduce the computational complexity of the beam training. If we use $\boldsymbol{F}_p$ and $\widetilde{\boldsymbol{v}}_p$ for combining, we have

$$\widetilde{\boldsymbol{z}}_p = \boldsymbol{F}_p \boldsymbol{h} x_p + \boldsymbol{F}_p \boldsymbol{\eta}, \quad (40)$$

and

$$\widetilde{y}_p = \widetilde{\boldsymbol{v}}_p \widetilde{\boldsymbol{z}}_p, \quad p = 1, 2, \ldots, QS + Q. \quad (41)$$

In fact, $\widetilde{\boldsymbol{z}}_p$ can not be obtained because we do not really perform beam training with $\boldsymbol{F}_p$. However, each entry of $\widetilde{\boldsymbol{z}}_p$ can be obtained from the beam training in (35) because both $\boldsymbol{F}_p$ and $\overline{\boldsymbol{W}}_m$ are composed of channel steering vectors from the same set $\boldsymbol{\Phi}$ in (28). We can obtain $\widetilde{\boldsymbol{z}}_p$ by setting

$$[\widetilde{\boldsymbol{z}}_p]_t = [\boldsymbol{z}_{\widetilde{m}_t}]_t, \quad (42)$$

where $\widetilde{m}_t$ can be obtained from (30) during the design of $\boldsymbol{F}_p$ in (38). Note that (41) can be computed in parallel to speed up the beam training. From $\{\widetilde{y}_1, \widetilde{y}_2, \ldots, \widetilde{y}_{QS+Q}\}$, we select one with the largest power, which can be expressed as

$$\widetilde{p} = \arg \max_{p=1,2,\ldots,QS+Q} |\widetilde{y}_p|^2. \quad (43)$$

From the hybrid-field codebook, $\boldsymbol{C}_{\text{h}}$, we select the codeword $[\boldsymbol{C}_{\text{h}}]_{:,\widetilde{p}}$ corresponding to the dedicated digital combiner $\widetilde{\boldsymbol{v}}_{\widetilde{p}}$ that can achieve the largest combining power $|\widetilde{y}_{\widetilde{p}}|^2$ as the result of the THBT. Moreover, if the user is in the near field, we can also roughly locate the user based on the beam training results. According to (37), the indices of the quantized angle and the quantized distance can be expressed as

$$\widetilde{q} = \lceil \widetilde{p}/S \rceil \text{ and } \widetilde{s} = \widetilde{p} - (\widetilde{q} - 1)S. \quad (44)$$

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

$$\widetilde{\Omega} = (2\widetilde{q} - 1 - Q)/Q \text{ and } \widetilde{r} = \frac{N^{3/2}\lambda S(1 - \Theta_{\widetilde{q}}^2)}{4\sqrt{2}\widetilde{s}}. \quad (45)$$

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

Compared to the hybrid-field beam sweeping based on $\boldsymbol{C}_{\text{h}}$ in (13) that needs $QS+Q$ times of beam training, the training overhead of the THBT scheme is substantially reduced to M. The computational complexity of the proposed scheme mainly comes from step 9 in **Algorithm 1** and is totally $\mathcal{O}(N_{\text{RF}}Q(S+1))$.

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

$$\widetilde{k} = -\frac{\lambda(1 - \widetilde{\Omega}^2)}{2\widetilde{r}} \text{ and } \widetilde{b} = \widetilde{\Omega} - \widetilde{k}(N + 1). \quad (46)$$

Note that the resolutions of $\widetilde{k}$ and $\widetilde{b}$ 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

$$\overline{\overline{\boldsymbol{W}}} = \text{blkdiag}\{\overline{\overline{\boldsymbol{w}}}_1^{\text{H}}, \overline{\overline{\boldsymbol{w}}}_2^{\text{H}}, \ldots, \overline{\overline{\boldsymbol{w}}}_{N_{\text{RF}}}^{\text{H}}\}, \quad (47)$$

where

$$[\overline{\overline{\boldsymbol{w}}}_t]_m = e^{j\pi(\widetilde{k}((t-1)M+m)^2 + \widetilde{b}((t-1)M+m))} \quad (48)$$

for $m = 1, 2, \ldots, M$ and $t = 1, 2, \ldots, N_{\text{RF}}$, so that the near-field effect can be alleviated. Note that $\overline{\overline{\boldsymbol{W}}}$ is designed based on $\widetilde{k}$ and $\widetilde{b}$. Then the received signal of the tth subarray after analog combining can be expressed as

$$\widehat{z}_t \overset{(a)}{\approx} \sum_{n=(t-1)M+1}^{tM} e^{-j\pi(\widetilde{k}n^2 + \widetilde{b}n)}[\boldsymbol{h}]_n$$

$$\overset{(b)}{\approx} g \sum_{n=(t-1)M+1}^{tM} e^{-j\pi(\widetilde{k}n^2 + \widetilde{b}n)}[\boldsymbol{\alpha}(N, \Omega, r)]_n$$

$$\overset{(c)}{\approx} \widetilde{g} \sum_{n=(t-1)M+1}^{tM} e^{-j\pi(\widetilde{k}n^2 + \widetilde{b}n)}[\boldsymbol{\gamma}(N, \Omega, r)]_n$$

$$= \widetilde{g} \sum_{n=(t-1)M+1}^{tM} e^{j\pi(\Delta k n^2 + \Delta b n)}$$

$$= \widetilde{g} \sum_{m=1}^{M} e^{j\pi(\Delta k(m+(t-1)M)^2 + \Delta b(m+(t-1)M))}$$

$$= \widetilde{g} e^{j\pi(\Delta k(t-1)^2 M^2 + \Delta b(t-1)M)}$$

$$\cdot \sum_{m=1}^{M} e^{j\pi(\Delta km^2 + \Delta bm + 2\Delta kmM(t-1))}, \quad (49)$$

for $t = 1, 2, \ldots, N_{\mathrm{RF}}$, where $\Delta k \triangleq k - \widetilde{k}$, $\Delta b \triangleq b - \widetilde{b}$ and $\widetilde{g} \triangleq g e^{j2\pi(\Omega(\delta_1 - 1/2) - \rho(N+1)^2/16)}$. In (49), 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 (49) is tricky due to the quadratic phase term Δkm^2. To obtain a deeper insight of $\widehat{z}_t$, we first focus on $e^{j\pi\Delta km^2}$. Define $\boldsymbol{s} \triangleq [e^{j\pi\Delta k}, e^{j\pi 4\Delta k}, \ldots, e^{j\pi M^2 \Delta k}]^{\mathrm{T}}$. According to the PSP [30]–[32], the discrete-time Fourier transform of $\boldsymbol{s}$ is expressed as

$$\mathcal{F}(\omega) = \sum_{m=1}^{M} [\boldsymbol{s}]_m e^{-j\pi m\omega}$$

$$\approx \begin{cases} e^{-j\pi(\frac{\omega^2}{4\Delta k} + \frac{1}{4})}\sqrt{\frac{1}{-\Delta k}}, & 2\Delta kM \leq \omega \leq 2\Delta k, \\ 0, & \text{others,} \end{cases}$$

$$(50)$$

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

$$[\boldsymbol{s}]_m = e^{j\pi\Delta km^2}$$

$$\approx \frac{1}{2}\int_{-1}^{1}\mathcal{F}(\omega)e^{j\pi m\omega}\mathrm{d}\omega$$

$$= \frac{1}{2}\int_{2\Delta kM}^{2\Delta k} e^{-j\pi(\frac{\omega^2}{4\Delta k} + \frac{1}{4})}\sqrt{\frac{1}{-\Delta k}}e^{j\pi m\omega}\mathrm{d}\omega. \quad (51)$$

Substituting (51) into (49), we have (52), which is shown at the top of the next page. In (52), we have

$$\overline{g} \triangleq \frac{1}{2\sqrt{-\Delta k}}\widetilde{g}e^{j\pi((M+1)\Delta b/2 - 1/4)},$$

$$\Delta\widetilde{k} \triangleq \Delta kM^2,$$

$$\Delta\widetilde{b} \triangleq (\Delta b + \Delta k(M+1))M,$$

$$\phi_t \triangleq \Delta b + 2\Delta kM(t-1). \quad (53)$$

Remark 1: From (52), the power of $\widehat{z}_t$ 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 w will not cancel each other. Then we have

$$|\phi_t + \omega| \leq \frac{1}{M}, \quad (54a)$$

$$\left|\frac{(M+1)\omega}{2} - \frac{\omega^2}{4\Delta k}\right| \leq 1/2, \quad (54b)$$

for $\omega \in [2\Delta kM, 2\Delta k]$. Substituting ϕ_t in (53) into (54a), we have

$$|\Delta k| \leq \frac{1}{M(N-1)} - \frac{1}{Q(N-1)}. \quad (55)$$

From (54b), we have

$$|\Delta k| \leq \frac{1}{4M^2}. \quad (56)$$

Note that

$$\frac{1}{M(N-1)} - \frac{1}{Q(N-1)} < \frac{1}{M(N-1)} \approx \frac{1}{N_{\mathrm{RF}}M^2} \leq \frac{1}{4M^2} \quad (57)$$

for $N_{\mathrm{RF}} \geq 4$, which is easy to be satisfied for XL-MIMO systems. Therefore, we omit the constraint in (56) and focus on (55) to simplify the expression. Note that

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

$$= \frac{\sqrt{2}}{2N^{3/2}S}. \quad (58)$$

Combining (56) and (58), we have

$$Q \geq M, \quad \text{and} \quad S \geq \frac{\sqrt{2}(N-1)}{2N^{3/2}\left(\frac{1}{M} - \frac{1}{Q}\right)}, \quad (59)$$

which provides guidance for the settings of S and Q.

In (52), the expression of $\widehat{z}_t$ is analytically derived, where the phase of $\widehat{z}_t$ shifts along with t. Specifically, the phase of $\widehat{z}_t$ is related to $\overline{g}$, $\mathcal{C}(t)$, and $\mathcal{B}(\phi_t)$, where $\overline{g}$ 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 $\mathcal{A}(\phi_t, \omega)$ is always positive under the constraint in (54a). Therefore, the unwrapped phase of $\widehat{z}_t$ is a quadratic function of t with ΔkM^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 $\Upsilon_t \triangleq \angle\widehat{z}_t$, where $\angle(\cdot)$ denotes the phase of a complex value. The first-order difference of Υ_t can be expressed as

$$\Delta\Upsilon_t \triangleq \Upsilon_{t+1} - \Upsilon_t = \pi(\Delta\widetilde{k}(2t-1) + \Delta\widetilde{b}), \quad (60)$$

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

$$\Delta^2\Upsilon_t \triangleq \Delta\Upsilon_{t+1} - \Delta\Upsilon_t = 2\pi\Delta\widetilde{k}, \quad (61)$$

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

$$|\Delta^2\Upsilon_t| = 2\pi|\Delta\widetilde{k}| \overset{(a)}{<} \frac{2\pi M^2}{M(N-1)} = \frac{2\pi}{N_{\mathrm{RF}}} \cdot \frac{N}{N-1} \overset{(b)}{<} \pi \quad (62)$$

where (a) can be obtained via (55) and (b) holds for $N_{\mathrm{RF}} > 2$. To avoid phase wrap, we can shift $\Delta^2\Upsilon_t$ to $[-\pi, \pi]$ by

$$\overline{\Upsilon}_t \triangleq \mathrm{mod}(\Delta^2\Upsilon_t + \pi, 2\pi) - \pi. \quad (63)$$

Then the estimation of Δk can be expressed as

$$\Delta\widehat{k} = \frac{1}{N_{\mathrm{RF}} - 2}\sum_{t=1}^{N_{\mathrm{RF}}-2}\frac{\overline{\Upsilon}_t}{2\pi M^2}. \quad (64)$$

Now we turn to the estimation of Δb. Define

$$\widetilde{\Upsilon}_t \triangleq \Delta\Upsilon_t - (2t-1)M^2\Delta\widehat{k}\pi - M(M+1)\Delta\widehat{k}\pi. \quad (65)$$

Then according to (60), we have

$$\widetilde{\Upsilon}_t = M\Delta b\pi + 2u\pi, \ u \in \mathbb{N}. \quad (66)$$

$$\widehat{z}_t \approx \widetilde{g}e^{j\pi(\Delta k(t-1)^2 M^2 + \Delta b(t-1)M)} \cdot \sum_{m=1}^{M} \left(\frac{1}{2} \int_{2\Delta kM}^{2\Delta k} e^{-j\pi\left(\frac{\omega^2}{4\Delta k} + \frac{1}{4}\right)} \sqrt{\frac{1}{-\Delta k}} e^{j\pi m\omega} d\omega \right) e^{j\pi(\Delta bm + 2\Delta kmM(t-1))}$$

$$= \widetilde{g}e^{j\pi(\Delta k(t-1)^2 M^2 + \Delta b(t-1)M)} \frac{1}{2} \int_{2\Delta kM}^{2\Delta k} e^{-j\pi\left(\frac{\omega^2}{4\Delta k} + \frac{1}{4}\right)} \sqrt{\frac{1}{-\Delta k}} \left(\sum_{m=1}^{M} e^{j\pi(\Delta bm + 2\Delta kmM(t-1))} e^{j\pi m w} \right) d\omega$$

$$= \widetilde{g}e^{j\pi(\Delta k(t-1)^2 M^2 + \Delta b(t-1)M)} \frac{1}{2} \sqrt{\frac{1}{-\Delta k}} \int_{2\Delta kM}^{2\Delta k} e^{-j\pi\left(\frac{\omega^2}{4\Delta k} + \frac{1}{4}\right)} \left(e^{j(M+1)(\pi\phi_t + \pi\omega)/2} \frac{\sin(M(\pi\phi_t + \pi\omega)/2)}{\sin((\pi\phi_t + \pi\omega)/2)} \right) d\omega$$

$$= \overline{g}e^{j\pi(\Delta \widetilde{k}(t-1)^2 + \Delta \widetilde{b}(t-1))} \int_{2\Delta kM}^{2\Delta k} \underbrace{e^{j\pi\left(\frac{(M+1)\omega}{2} - \frac{\omega^2}{4\Delta k}\right)}}_{\mathcal{P}(\omega)} \underbrace{\frac{\sin(M(\pi\phi_t + \pi\omega)/2)}{\sin((\pi\phi_t + \pi\omega)/2)}}_{\mathcal{A}(\phi_t, \omega)} d\omega$$

$$= \overline{g}\underbrace{e^{j\pi(\Delta \widetilde{k}(t-1)^2 + \Delta \widetilde{b}(t-1))}}_{\mathcal{C}(t)} \underbrace{\int_{2\Delta kM}^{2\Delta k} \mathcal{P}(\omega)\mathcal{A}(\phi_t, \omega)d\omega}_{\mathcal{B}(\phi_t)}$$

$$= \overline{g}\mathcal{C}(t)\mathcal{B}(\phi_t). \tag{52}$$

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

---

1: **Input:** N, N_{RF}, M, S, λ, $\overline{z}_t$, $\widetilde{\Omega}$ and $\widetilde{r}$.
2: Obtain $\widetilde{k}$ and $\widetilde{b}$ via (46).
3: Obtain $\Upsilon_t = \angle \overline{z}_t$, for $t = 1, 2, \ldots, N_{\mathrm{RF}}$.
4: Obtain $\Delta \Upsilon_t$ and $\Delta^2 \Upsilon_t$ via (60) and (61), respectively.
5: Obtain $\Delta \widehat{k}$ and $\Delta \widehat{b}$ via (64) and (68), respectively.
6: Obtain $\widehat{k}$ and $\widehat{b}$ via (69).
7: Obtain $\widehat{\Omega}$ and $\widehat{r}$ via (70).
8: **Output:** $\widehat{k}$, $\widehat{b}$, $\widehat{\Omega}$ and $\widehat{r}$.

Note that

$$|M\Delta b\pi| \leq \frac{M\pi}{Q} + \frac{(N+1)\pi}{(N-1)} - \frac{M(N+1)\pi}{Q(N-1)} \overset{(a)}{\approx} \pi. \tag{67}$$

where (a) holds because $(N+1)/(N-1) \approx 1$ for XL-MIMO systems. Therefore, we can also shift the results in (66) to $[-\pi, \pi]$ to avoid phase wrap. Define $\widehat{\Upsilon}_t \triangleq \mathrm{mod}(\Upsilon_t + \pi, 2\pi) - \pi$. Then the estimation of Δb can be expressed as

$$\Delta \widehat{b} = \frac{1}{N_{\mathrm{RF}} - 1} \sum_{t=1}^{N_{\mathrm{RF}}-1} \frac{\widehat{\Upsilon}_t}{M\pi}. \tag{68}$$

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

$$\widehat{k} = \Delta \widehat{k} + \widetilde{k} \text{ and } \widehat{b} = \Delta \widehat{b} + \widetilde{b}. \tag{69}$$

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

$$\widehat{\Omega} = \widehat{b} + \widehat{k}(N+1) \text{ and } \widehat{r} = -\frac{\lambda(1 - \widehat{\Omega}^2)}{2\widehat{k}}. \tag{70}$$

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

Remark 2: The overheads, computation 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 computation complexity. The computation of the BRPSS mainly comes from the computing of (60), (61), (64) and (68), which are all only related to N_{RF}. Therefore, the computation complexity of the BRPSS is $\mathcal{O}(N_{\mathrm{RF}})$.

Note that the unwrapped phase of $\widehat{z}_t$ is the quadratic function of the indices of subarrays. Therefore, at least 3 subarrays are needed for the BRPSS scheme.

Remark 3: The BRPSS is developed based on the phase shifts of the subarrays. Therefore, it can also be extended to the fully-connected structure if the overlapped or non-overlapped subarrays can be constructed by deactivating part of antennas.

V. NEAR-FIELD BEAM TRACKING

In this section, we propose a low-complexity near-field beam tracking (NFBT) scheme, where a kinematic model is adopted to characterize the channel variations at different time slots 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.

Note that the near-field channel is related to both the angle and the distance of the user, which facilitates the implementation of beam tracking. Specifically, based on the estimated angle and distance of the user, the position of the user can be calculated. Therefore, the kinematic model can be exploited to characterize the channel variations at different time slots. In Fig. 3, 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
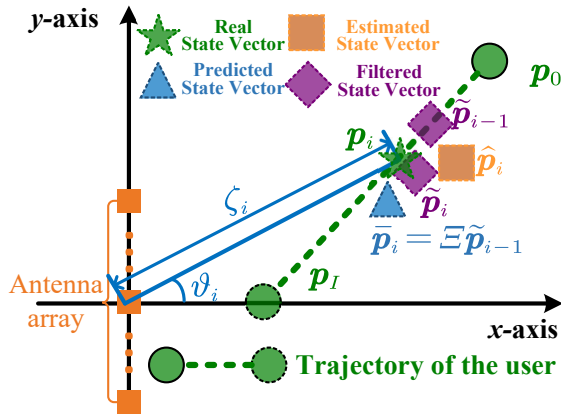
Fig. 3.   Illustration of the near-field beam tracking procedure.

Algorithm 3 Near-Field Beam Tracking (NFBT) Scheme

1: **Input:** N, N_{RF}, M, S, λ, $\boldsymbol{p}[0]$ and ΔT.
2: **Initialization:** $i \leftarrow 1$.
3: **while** stop conditions in Sec. V-5 are not satisfied **do**
4:     Obtain $\overline{\boldsymbol{p}}[i]$ via (71).
5:     Obtain $\widehat{\boldsymbol{p}}[i]$ via (77).
6:     Obtain $\widetilde{\boldsymbol{p}}[i]$ via the EKF.
7:     Obtain $\boldsymbol{h}_i$ via (78).
8:     $i \leftarrow i + 1$.
9: **end while**
10: **Output:** $\widetilde{\boldsymbol{h}}_i$.

---

the BS in a block-wise way, where the duration of one block is ΔT. Denote the kinematic parameters of the user at the ith block as $\boldsymbol{p}[i] = [a_x[i], a_y[i], v_x[i], v_y[i]]^{\mathrm{T}}$, 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 ith 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 $\boldsymbol{p}[0] = [a_x[0], a_y[0], 0, 0]^{\mathrm{T}}$, which can be obtained via the THBT and BRPSS schemes.

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

$$\overline{\boldsymbol{p}}[i] = \underbrace{\begin{bmatrix} 1 & 0 & \Delta T & 0 \\ 0 & 1 & 0 & \Delta T \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}}_{\Xi} \widetilde{\boldsymbol{p}}[i-1] \qquad (71)$$

where the predicted distance and angle can be expressed as

$$\overline{\zeta}_i = \sqrt{\overline{a}_x[i]^2 + \overline{a}_y[i]^2} \text{ and } \overline{\vartheta}_i = \arctan(\overline{a}_y[i]/\overline{a}_x[i]). \quad (72)$$

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

$$\overrightarrow{\boldsymbol{W}}_i = \mathrm{blkdiag}\{\overrightarrow{\boldsymbol{w}}_1^{\mathrm{H}}, \overrightarrow{\boldsymbol{w}}_2^{\mathrm{H}}, \ldots, \overrightarrow{\boldsymbol{w}}_{N_{\mathrm{RF}}}^{\mathrm{H}}\}, \qquad (73)$$

where

$$[\overrightarrow{\boldsymbol{w}}_t]_m = e^{j\pi(\overline{k}_i((t-1)M+m)^2 + \overline{b}_i((t-1)M+m))} \qquad (74)$$

for $m = 1, \ldots, M$ and $t = 1, 2, \ldots, N_{\mathrm{RF}}$. In (74),

$$\overline{k} = -\frac{\lambda(1 - \sin^2 \overline{\vartheta}_i)}{2\overline{\zeta}_i} \text{ and } \overline{b} = -\sin \overline{\vartheta}_i - \overline{k}(N+1) \quad (75)$$

which are calculated based on the predicted distance and angle in (72). Then the received signal of the tth subarray after analog combining can be expressed as

$$\overline{z}_t = \overrightarrow{\boldsymbol{w}}_t^{\mathrm{H}} \boldsymbol{G}_t \overline{\boldsymbol{h}}_i + \overrightarrow{\boldsymbol{w}}_t^{\mathrm{H}} \boldsymbol{G}_t \overline{\boldsymbol{\eta}}_i, \qquad (76)$$

for $t = 1, 2, \ldots, N_{\mathrm{RF}}$, where $\overline{\boldsymbol{h}}_i$ denotes the channel between the user and the BS at the ith block and $\overline{\boldsymbol{\eta}}_i$ denotes the AWGN. Substituting N, N_{RF}, M, S, λ, $\overline{z}_t$, $-\sin \overline{\vartheta}_i$ and $\overline{\zeta}_i$ into **Algorithm 2**, we can obtain estimated angle $\widehat{\vartheta}_i$ and estimated distance $\widehat{\zeta}_i$ of the user at the ith block. Then the estimated state vector can be expressed as

$$\widehat{\boldsymbol{p}}[i] = \left[\widehat{\zeta}_i \cos \widehat{\vartheta}_i, \widehat{\zeta}_i \sin \widehat{\vartheta}_i, 0, 0\right]^{\mathrm{T}}. \qquad (77)$$

4) Filtering: The estimated stated vector, $\widehat{\boldsymbol{p}}[i]$, and the predicted state vector, $\overline{\boldsymbol{p}}[i]$, can be exploited by the widely-adopted EKF for effective tracking. Since the EKF has been exhaustively introduced in a variety of works [33]–[37], we omit the details and denote the filtered state vector at the ith block after performing the EKF as $\widetilde{\boldsymbol{p}}[i]$. Then the estimated LoS channel at the ith block can be expressed as

$$[\widetilde{\boldsymbol{h}}_i]_n = e^{j\pi\left(\overrightarrow{k} n^2 + \overrightarrow{b} n\right)}$$

$$\overrightarrow{k} = -\frac{\lambda(1 - \sin^2 \widetilde{\vartheta}_i)}{2\widetilde{\zeta}_i}, \quad \overrightarrow{b} = -\sin \widetilde{\vartheta}_i - \overrightarrow{k}(N+1),$$

$$\widetilde{\zeta}_i = \sqrt{\widetilde{a}_x[i]^2 + \widetilde{a}_y[i]^2}, \text{ and } \widetilde{\vartheta}_i = \arctan(\widetilde{a}_y[i]/\widetilde{a}_x[i]), \quad (78)$$

for $n = 1, 2, \ldots, N$, which can be used to design the hybrid precoders for the 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.

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

VI. SIMULATION RESULTS

Now we evaluate the performance of the proposed schemes. We consider an XL-MIMO system equipped with $N = 512$ antennas. The antenna array is composed of $N_{\mathrm{RF}} = 4$ subarrays with each subarray having $M = 128$ antennas. The wavelength is set to be $\lambda = 0.003$ m corresponding to the carrier frequency of 100 GHz. 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
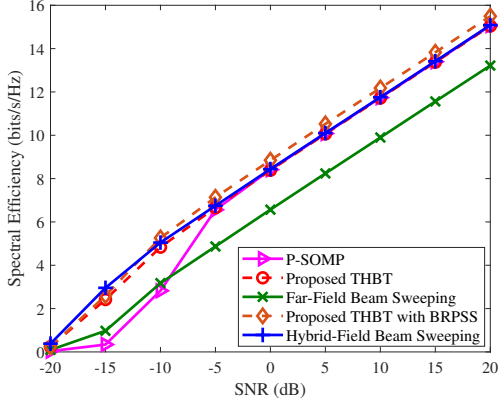
Fig. 4.   Comparisons of the spectral efficiency for different methods.

Fig. 5.   Comparisons of the beamforming gains for different methods.

Fig. 6.   Comparisons of the positioning error for different methods.

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 $g_3 \sim \mathcal{CN}(0,0.01)$. Channel angle Ω_l of the lth path obeys the uniform distribution between $[-\sqrt{3}/2, \sqrt{3}/2]$. To compare the proposed schemes with P-SOMP [13] fairly, we set the pilot length of P-SOMP to be $M = 128$, which is also the training overhead of the THBT.

A. Evaluation of the THBT and BRPSS Schemes

In Fig. 4, we compare the proposed schemes with P-SOMP, the far-filed beam sweeping [38] and the hybrid-field beam sweeping in terms of spectral efficiency for different SNRs. The distances between the BS and the user or scatterers obey the uniform distribution between $[10, 30]$ m. 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. 4, the hybrid-field beam sweeping can achieve better performance than the other four schemes when the SNR is less than -10 dB, which lies in the fact that hybrid-field beam sweeping exhaustively tests all the code-words in C_{h} and needs far more times of beam training than the other four 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 four methods 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. Furthermore, the performance of the proposed THBT scheme can approach the performance of the hybrid-field beam sweeping at various SNR conditions. Moreover, the performance of P-SOMP is worse than that of the other four schemes at low SNRs, such as -15 dB, because the random beamforming of P-SOMP cannot achieve enough beamforming gain and will significantly degrade the performance. The performance of the far-field beam sweeping is worse than that of the other four schemes at high SNRs because the beamforming gain of the far-field channel steering vector will decrease in the near field.

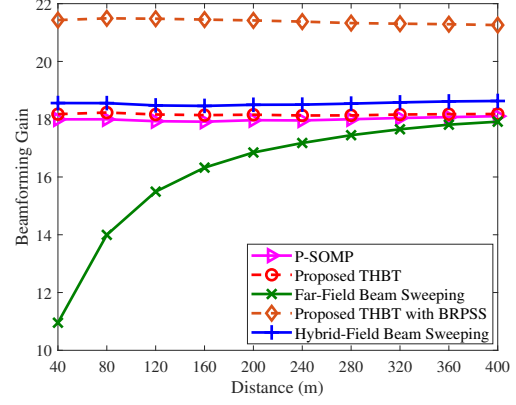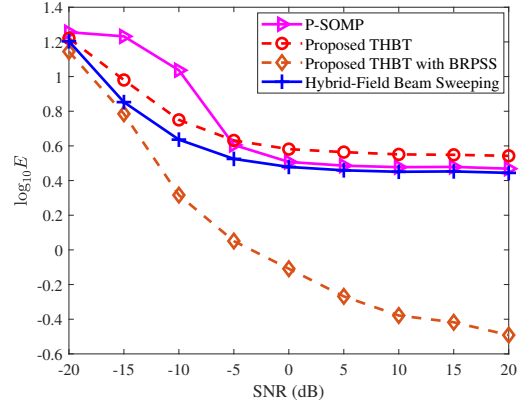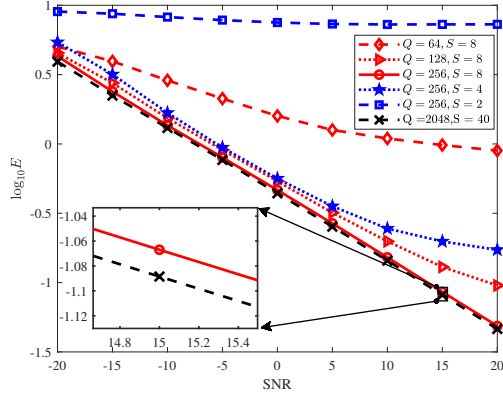In Fig. 5, we compare the proposed schemes with P-SOMP, the far-filed beam sweeping and the hybrid-field beam sweeping in terms of beamforming gain for different distances. The distances between the BS and the user or scatterers obey the uniform distribution between $[10, r]$ m, where r ranges from 40 to 400. The SNR is fixed to be -10 dB. From Fig. 5, 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 far-field beam sweeping will suffer severe loss of beamforming gain because it only considers the far-field channels. By contrast, the other four schemes are robust to the distance because they consider both the near-filed channel and the far-field channel. In addition, the performance of the proposed THBT scheme is better than P-SOMP and can approach that of the hybrid-field beam sweeping with only a slight loss of beamforming gain at different distances.

In Fig. 6, we evaluate the positioning performance of the proposed schemes. Since the far-field beam sweeping cannot obtain the position of the user, we only compare the proposed schemes with P-SOMP and the hybrid-filed beam sweeping. We denote the distance between the real position and the estimated position as E. From Fig. 6, 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,

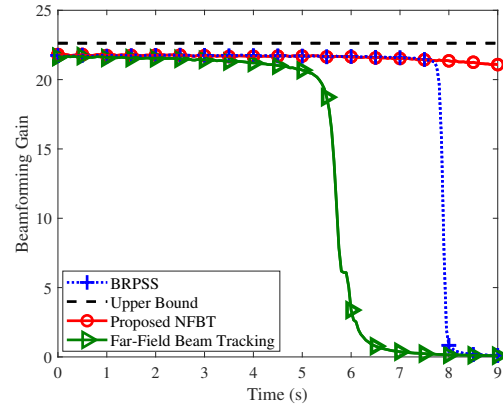Fig. 7. Evaluation of the BRPSS for different settings of Q and S.

Fig. 8. Comparisons of the beamforming gains during the tracking process.

Fig. 9. Comparisons of the averaged spectral efficiency during the tracking process.

the proposed THBT performs better than P-SOMP when SNR is less than -5 dB because the random beamforming of P-SOMP cannot achieve enough beamforming gain and significantly degrade the performance. Moreover, P-SOMP performs better than THBT when SNR is larger than -5 dB because P-SOMP exploits the measures of all the pilots and will suffer from high computation complexity while THBT only uses N_{RF} of the measures to form the directional beams.

We also compare the training overheads of different schemes in Fig. 4 to 6. The training overheads of the hybrid-field beam sweeping, the far-field beam sweeping, P-SOMP, the proposed THBT, and the proposed THBT with the BRPSS are $Q(S + 1)$, N, M, M, and $M + 1$, respectively. Under the simulation settings, these five schemes require 6144, 512, 128, 128, and 129 time slots, respectively, where the proposed THBT can approach the performance of the hybrid-field beam sweeping with 97.92% reduction in training overhead, and the proposed THBT with the BRPSS outperforms the hybrid-field beam sweeping with 97.90% reduction in training overhead.

In Fig. 7, we evaluate the performance of the BRPSS for different settings of Q and S. To remove the effects of the NLoS paths and the prior beam training failure, we set $L = 1$ and suppose the beam training has found the codeword best fitting for the channel in the hybrid-field codebook. From Fig. 7, the positioning error decreases with the increase of S or Q if we fix $Q = 256$ or $S = 8$. This is because beam gain losses of subarrays will occur if the conditions of Q and S in (59) are not satisfied, which will deteriorate the performance of beam refinement. However, if the conditions of Q and S in (59) 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 (59) 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 $\boldsymbol{p}_0 = [50, 50\sqrt{3}, 0, 0]$, $\Delta T = 0.05$ s, $I = 180$ and $[v_x[i], v_y[i]]^{\mathrm{T}} = [-5, -5\sqrt{3}]^{\mathrm{T}}$ for $i = 1, 2, \ldots, I$.

In Fig. 8, we compare the proposed NFBT with the BRPSS and the far-field beam tracking [26] in terms of the beam-

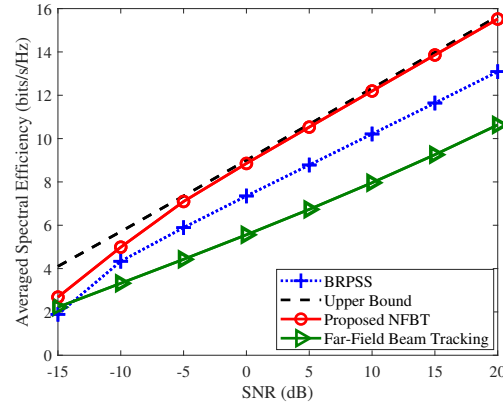forming gain during the tracking process, where the BRPSS is performed based on the estimation of the previous block. The SNR is set to be 0 dB. From the figure, the performance of the BRPSS drops dramatically at the 8th second because the deviation between the positions of the adjacent blocks is too large and the deteriorated beamforming gains cannot support effective beam tracking. In addition, the performance of the far-field beam tracking 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. 9, we compare the proposed NFBT with the BRPSS and the far-field beam tracking in terms of the averaged spectral efficiency during the tracking process for different SNRs. From the figure, the proposed NFBT performs the best among the three schemes and can approach the upper bound when SNR is larger than -5dB, which demonstrates

the effectiveness of the proposed schemes.

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 slots and the EKF has been exploited for beam tracking. Future works will be continued with effective beam training and high-resolution beam refinement for fully-connected structures.

APPENDIX A

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

$$
\begin{aligned}
|G(\boldsymbol{G}_t\boldsymbol{h}, \Omega)| &= \sqrt{M}\boldsymbol{\beta}(M, \Omega)^{\mathrm{H}}\boldsymbol{G}_t\boldsymbol{h} \\
&\approx \left| \sum_{n=(t-1)M+1}^{tM} e^{j\pi(kn^2+bn)}e^{-j\pi n\Omega} \right| \\
&= \left| \sum_{n=1}^{M} e^{j\pi(kn^2+\bar{\bar{b}}n)}e^{-j\pi n\Omega} \right| \\
&\approx \left| \int_1^M e^{j\pi(k\chi^2+(\bar{\bar{b}}-\Omega)\chi))}\mathrm{d}\chi \right| \\
&= \left| \int_{-\infty}^{\infty} U(\chi)e^{j\pi(k\chi^2+(\bar{\bar{b}}-\Omega)\chi))}\mathrm{d}\chi \right|, \quad (79)
\end{aligned}
$$

where

$$
\bar{\bar{b}} \triangleq b + 2k(t-1)M, \quad U(\chi) = \begin{cases} 1, & 1 \le \chi \le M, \\ 0, & \text{others.} \end{cases} \quad (80)
$$

According to the PSP [30]–[32], (79) can be approximated as

$$
\begin{aligned}
|G(\boldsymbol{G}_t\boldsymbol{h}, \Omega)| &\approx \left| \sqrt{\frac{-2\pi}{\Phi''(\chi_0, \Omega)}}U(\chi_0) \right| \\
&= \begin{cases} \sqrt{\frac{1}{-k}}, & \bar{\bar{b}}+2kM \le \Omega \le \bar{\bar{b}}+2k, \\ 0, & \text{others,} \end{cases} \quad (81)
\end{aligned}
$$

where $\Phi(\chi, \Omega) \triangleq \pi(k\chi^2 + (\bar{\bar{b}} - \Omega)\chi))$ and $\chi_0 = \frac{\Omega-\bar{\bar{b}}}{2k}$. The relation in (81) 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 calculated as

$$
\begin{aligned}
\Gamma &= 1 - \frac{\max_{\Omega} |G(\boldsymbol{G}_t\boldsymbol{h}, \Omega)|}{M} \\
&= 1 - \sqrt{\frac{4r}{\lambda(1 - \Omega^2)M^2}} \\
&\overset{(a)}{\le} 1 - \sqrt{\frac{4r}{\lambda M^2}} \\
&\overset{(b)}{\le} 1 - \sqrt{\frac{2\sqrt{D^3/\lambda}}{\lambda M^2}} \\
&= 1 - \frac{N_{\mathrm{RF}}}{\sqrt[4]{2N}}, \quad (82)
\end{aligned}
$$

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

$$
\Gamma_{\max} = \max\{1 - N_{\mathrm{RF}}/\sqrt[4]{2N}, 0\}, \quad (83)
$$

which completes the proof.

APPENDIX B

First of all, we calculate the beam center of $\boldsymbol{G}_t\boldsymbol{\alpha}(N, \Omega, r)$. Based on (81), we can approximate the beam gain of $\boldsymbol{G}_t\boldsymbol{\alpha}(N, \Omega, r)$ as

$$
|G(\boldsymbol{G}_t\boldsymbol{\alpha}(N, \Omega, r), \Omega)| = \begin{cases} \sqrt{\frac{1}{-k}}, & \bar{\bar{b}}+2kM \le \Omega \le \bar{\bar{b}}+2k, \\ 0, & \text{others,} \end{cases} \quad (84)
$$

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

$$
\begin{aligned}
B_t &= \bar{\bar{b}} + k(M + 1) \\
&= \Omega + \frac{\lambda(1 - \Omega^2)(N - (2t - 1)M)}{4r}. \quad (85)
\end{aligned}
$$

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

$$
\begin{aligned}
\Psi_t &= \frac{r\Omega - \Delta_t\lambda}{\sqrt{r^2 + \Delta_t^2\lambda^2 - 2r\Omega\Delta_t\lambda}} \\
&= \frac{r\Omega - \Omega^2\Delta_t\lambda}{\sqrt{r^2 + \Delta_t^2\lambda^2 - 2r\Omega\Delta_t\lambda}} + \frac{(\Omega^2 - 1)\Delta_t\lambda}{\sqrt{r^2 + \Delta_t^2\lambda^2 - 2r\Omega\Delta_t\lambda}} \\
&\overset{(a)}{\approx} \frac{r\Omega - \Omega^2\Delta_t\lambda}{r - \Omega\Delta_t\lambda} + \frac{(\Omega^2 - 1)\Delta_t\lambda}{r} \\
&= \Omega + \frac{\lambda(1 - \Omega^2)(N - (2t - 1)M)}{4r}, \quad (86)
\end{aligned}
$$

where the former of (a) holds because we approximate $\sqrt{r^2 + \Delta_t^2\lambda^2 - 2r\Omega\Delta_t\lambda}$ with its one-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 (86) 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 (85) and (86), we have

$$B_t \approx \Psi_t, \tag{87}$$

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] E. Bjornson, L. Van der Perre, S. Buzzi, and E. G. Larsson, "Massive MIMO in sub-6 GHz and mmWave: Physical, practical, and use-case differences," *IEEE Wireless Commun.*, vol. 26, no. 2, pp. 100–108, Apr. 2019.

[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] Z. Qiu, S. Zhou, M. Zhao, and W. Zhou, "Low-complexity precoding by exploiting spatial sparsity in massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 21, no. 7, pp. 4740–4753, Dec. 2022.

[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., early access*, pp. 1–71, 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. 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.

[8] 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.

[9] M. Cui, Z. Wu, Y. Lu, X. Wei, and L. Dai, "Near-field MIMO communications for 6G: Fundamentals, challenges, potentials, and future directions," *IEEE Commun. Mag.*, vol. 61, no. 1, pp. 40–46, Feb. 2023.

[10] C. Feng, W. Shen, J. An, and L. Hanzo, "Weighted sum rate maximization of the mmwave cell-free MIMO downlink relying on hybrid precoding," *IEEE Trans. Wireless Commun.*, vol. 21, no. 4, pp. 2547–2560, Apr. 2022.

[11] X. Sun, C. Qi, and G. Y. Li, "Beam training and allocation for multiuser millimeter wave massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 18, no. 2, pp. 1041–1053, Feb. 2019.

[12] S. Srivastava, R. K. Singh, A. K. Jagannatham, A. Chockalingam, and L. Hanzo, "OTFS transceiver design and sparse doubly-selective CSI estimation in analog and hybrid beamforming aided mmWave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 21, no. 12, pp. 10 902–10 917, July 2022.

[13] 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.

[14] 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, Sep. 2017.

[15] 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.

[16] Z. Zhang, X. Wu, and D. Liu, "Joint precoding and combining design for hybrid beamforming systems with subconnected structure," *IEEE Syst. J.*, vol. 14, no. 1, pp. 184–195, Mar. 2020.

[17] 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.

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

[19] 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.

[20] 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.

[21] 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.

[22] X. Shi, J. Wang, Z. Sun, and J. Song, "Hierarchical codebook-based beam training for extremely large-scale massive MIMO," *arXiv preprint arXiv:2210.03345*, 2022.

[23] 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.

[24] J. A. Zhang, K. Wu, X. Huang, and Y. J. Guo, "Beam alignment for analog arrays based on gaussian approximation," *IEEE Trans. Veh. Technol., early access*, pp. 1–6, 2023.

[25] 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.

[26] 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.

[27] 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.

[28] 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.

[29] 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.

[30] C. M. Bender and S. A. Orszag, *Advanced Mathematical Methods for Scientists and Engineers*.   New York, NY, USA: McGraw-Hill, 1978.

[31] F. W. J. Olver, *Asymptotics and Special Functions*.   Boca Raton, FL, USA: CRC Press, 1997.

[32] N. M. Temme, *Asymptotic Methods for Integrals*.   Singapore: World Scientific, vol. 11, 2014.

[33] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*.   Cliffs, NJ, USA: Prentice-Hall, 1998.

[34] M. A. Richards, *Fundamentals of Radar Signal Processing*, 2nd ed. New York, NY, USA: McGraw-Hill, 2005.

[35] 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.

[36] 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.

[37] H. Wang, R. Lu, Z. Peng, and M. Li, "Timestamp-free clock parameters tracking using extended Kalman filtering in wireless sensor networks," *IEEE Trans. Commun.*, vol. 69, no. 10, pp. 6926–6938, Oct. 2021.

[38] 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.

Beam Training and Tracking for Extremely Large-Scale MIMO mmWave/THz 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 millimeter wave (mmWave) and terahertz (THz) communication systems, where extremely large-scale multiple-input-multiple-output (MIMO) with partially-connected hybrid combining structures is adopted. First, we propose a two-stage hybrid-field beam training scheme for both the near field and 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

Millimeter wave (MmWave) and terahertz (THz) have attracted great interests due to their abundant spectrum resources. In the existing fifth generation wireless communications, massive multiple-input-multiple-output (MIMO) has been integrated into the mmWave communications to improve the spectral efficiency by exploiting the spatial degree of freedom [2]–[4]. For future sixth generation wireless communications, extremely large-scale MIMO (XL-MIMO) with far more antennas than the existing massive MIMO is considered [5]–[9].

Due to large power consumption, the fully-digital structure that allocates each antenna with a dedicated radio frequency (RF) chain is impractical for large antenna arrays [10]–[12]. Consequently, the hybrid structure, where a small number of RF chains are connected to a large number of antennas, is developed for XL-MIMO systems [13]. According to the ways how the RF chains are connected to the antennas, the hybrid structure can generally be divided into two categories, including fully-connected structure and partially-connected structure. Although the fully-connected structure can achieve better spectral efficiency than the partially-connected 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 [14]–[16].

One important difference between XL-MIMO and the existing massive MIMO is the channel features. Depending on the distance between the user and the BS, the radiation field can be divided into the near field and the far field, bounded by the Rayleigh distance [17]–[19]. 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 systems, especially when a user is close to the BS. In this context, channel state information (CSI) acquisition methods for XL-MIMO should take both the near-field and far-field effects into consideration.

In general, CSI acquisition includes channel estimation and beam training [20]. Channel estimation usually focuses on efficient estimation of the high-dimension channel matrix by

exploiting advanced signal processing techniques, such as compressed sensing, while the beam training can avoid the estimation of the high-dimension channel matrix and obtain considerable beamforming gain, especially in low signal-to-noise ratio (SNR). In [13], 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. In [21], hierarchical codebooks for both the far field and the near field are designed based on the uniformly quantized space. In [22], a chirp-based hierarchical codebook is designed based on the expressions of the near-field channels in the slope-intercept domain.

One common challenge on the 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 for angle estimation of far-field channels [23]–[25]. In [23], beam refinement with a closed-form solution based on the monopulse signals is developed. In [24], an efficient angle-of-arrival estimator is designed by approximating the power of the array response as a Gaussian function. In [25], an auxiliary beam pair is designed to provide high-resolution estimate for the angle-of-departure of the channel. These 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 heavy burden of training overheads if still using existing methods. An efficient way to reduce the training overhead is the beam tracking, which exploits the correlations of the channels at different time slots to narrow down the sets of training candidates. There are a variety of beam tracking schemes for the far-field channels [26]–[29]. In [26], well-designed pairs of auxiliary beams can capture the angle variations. The beam zooming-based beam tracking scheme in [27] exploits the delay-phase precoding structure to flexibly control the angular coverage of frequency-dependent beams. In the adaptive tracking framework in [28], beam direction is updated according to measurements of the current data beam. The grid-based hybrid tracking scheme in [29] searches the surroundings of the former beam and selects the best beam according to the changing trend of the previously used beams.

To the best knowledge of the authors, so far there has been no literature working on effective beam training, refinement, and tracking for XL-MIMO systems with near-field effect. To fill in the gaps, we investigate these issues for XL-MIMO systems with partially-connected structures in this paper. 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 a subarray 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, channel parameters can be estimated with high resolution.

- 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, a 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. $(\cdot)^{\mathrm{T}}$ and $(\cdot)^{\mathrm{H}}$ denote the transpose and conjugate transpose (Hermitian), respectively. $[\boldsymbol{A}]_{:,m}$ denotes the mth column of a matrix $\boldsymbol{A}$. 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 $\mathrm{blkdiag}\{\cdot\}$ denote the floor operation and the block diagonalization operation, respectively. Moreover, $\mathcal{O}$ and $\mathrm{mod}(\cdot)$ denote the order
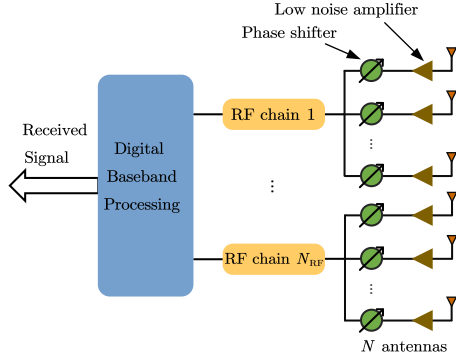
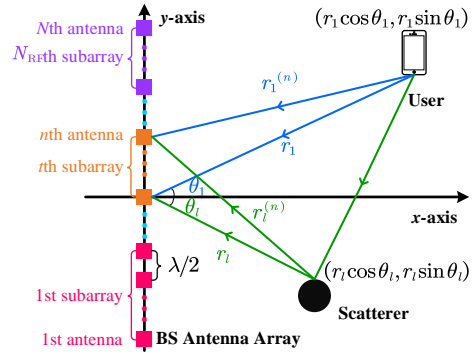Fig. 1: Illustration of a partially-connected hybrid combining structure.

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

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 hybrid structure has far more antennas than RF chains, i.e., $N \gg N_{\mathrm{RF}}$. The ULA is formed by N_{RF} non-overlapping subarrays, where each subarray has $M = N/N_{\mathrm{RF}}$ 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 analog combining and digital combining at the BS side while a single-antenna user is considered for simplification.

During uplink beam training, the pth transmit signal by the user is denoted as x_p for $p = 1, 2, \ldots, P$, where P is the signal length. The channel vector between the user and the BS is denoted as $\boldsymbol{h} \in \mathbb{C}^N$. Then the signal after the analog combining at the BS side can be expressed as [15]

$$y_p = \boldsymbol{v}_p \boldsymbol{W}_p \boldsymbol{h} x_p + \boldsymbol{v}_p \boldsymbol{W}_p \boldsymbol{\eta}, \tag{1}$$

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

Generally, two kinds of channels, including 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 [17]–[19]. The commonly used boundary distance to distinguish the near field and the far field is the Rayleigh distance

$$Z = 2D^2/\lambda, \tag{2}$$

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-filed channel; otherwise, it is the near-filed 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 $Z = N\lambda^2/2$.

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 nth antenna is $(0, \delta_n\lambda)$, where $\delta_n \triangleq (2n - N - 1)/4$ for $n = 1, 2, \ldots, N$. The coordinate of the user is denoted as $\boldsymbol{p}_1 = (r_1 \cos\theta_1, r_1 \sin\theta_1)$, 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 lth path for $l \geq 2$ is denoted as $\boldsymbol{p}_l = (r_l \cos\theta_l, r_l \sin\theta_l)$, where r_l is the distance between the scatterer on the lth path and the coordinate origin, and $\theta_l \in [-\pi/2, \pi/2]$ is the angle of the scatterer on the lth path relative to the positive x-axis. The distance between $\boldsymbol{p}_l$ and the nth antenna can be expressed as

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

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

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

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

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

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 simplified as $r_l^{(n)} \approx r_l - \Omega_l \delta_n \lambda$ because $\delta_n \lambda / r_l \approx 0$ and $\sqrt{1+\epsilon} \approx 1 + \frac{1}{2}\epsilon$ [13]. As a result, $\boldsymbol{\alpha}(N, \Omega_l, r_l)$ in (5) can be approximated as

$$\boldsymbol{\alpha}(N, \Omega_l, r_l) \approx \frac{1}{\sqrt{N}} \left[e^{j2\pi\Omega_l\delta_1}, \ldots, e^{j2\pi\Omega_l\delta_N} \right]^{\mathrm{T}}. \tag{6}$$

Removing a constant factor $e^{j2\pi\Omega_l\delta_1}$ from all the entries of $\boldsymbol{\alpha}(N, \Omega_l, r_l)$ in (6), we have

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

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

If $r_l \le Z$, the approximation in (7) is not accurate enough. Alternatively, we simplify $r_l^{(n)}$ 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 $r_l^{(n)} \ge 0.5\sqrt{D^3/\lambda}$ [13], [18]. Note that the ratio of Z to $0.5\sqrt{D^3/\lambda}$ is $4\sqrt{N}$, which indicates that the latter is much smaller than the former, especially for XL-MIMO systems. For example, if $N = 1024$, $0.5\sqrt{D^3/\lambda}$ is only $1/128$ of Z, which is almost negligible. Therefore, in this work, we focus on the near field with $r_l^{(n)} \ge 0.5\sqrt{D^3/\lambda}$. Then $\boldsymbol{\alpha}(N, \Omega_l, r_l)$ in (5) can be approximated as

$$\boldsymbol{\alpha}(N, \Omega_l, r_l) \approx \frac{1}{\sqrt{N}} \left[e^{j2\pi(\Omega_l\delta_1 - \rho_l\delta_1^2)}, \ldots, e^{j2\pi(\Omega_l\delta_N - \rho_l\delta_N^2)} \right]^{\mathrm{T}} \tag{8}$$

where $\rho_l \triangleq \frac{\lambda(1-\Omega_l^2)}{2r_l}$. Define $b_l \triangleq \Omega_l + \rho_l(N+1)/2$ and $k_l \triangleq -\rho_l/2$. Removing a constant factor $e^{j2\pi(\Omega_l(\delta_1-1/2) - \rho_l(N+1)^2/16)}$ from all the entries of $\boldsymbol{\alpha}(N, \Omega_l, r_l)$ in (6), we have

$$e^{-j2\pi(\Omega_l(\delta_1-1/2) - \rho_l(N+1)^2/16)} \boldsymbol{\alpha}(N, \Omega_l, r_l)$$

$$\approx \frac{1}{\sqrt{N}} \left[e^{j\pi(k_l+b_l)}, \ldots, e^{j\pi(k_l n^2 + b_l n)}, \ldots, e^{j\pi(k_l N^2 + b_l N)} \right]^{\mathrm{T}}$$

$$\triangleq \boldsymbol{\gamma}(N, \Omega_l, r_l). \tag{9}$$

Note that $\boldsymbol{\beta}(N, \Omega_l)$ is a special case of $\boldsymbol{\gamma}(N, \Omega_l, r_l)$ with $\rho_l = 0$. Therefore, $\boldsymbol{\gamma}(N, \Omega_l, r_l)$ 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 and far fields.

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 $\boldsymbol{C}_{\mathrm{f}} \in \mathbb{C}^{N \times Q}$ based on the far-field channel steering vectors [3], where Q denotes the quantized number of the angle and the qth column of $\boldsymbol{C}_{\mathrm{f}}$ is denoted as

$$[\boldsymbol{C}_{\mathrm{f}}]_{:,q} \triangleq \boldsymbol{\beta}(N, (2q - 1 - Q)/Q) \tag{10}$$

for $q = 1, 2, \ldots, Q$. Then we establish a near-field codebook $\boldsymbol{C}_{\mathrm{n}} \in \mathbb{C}^{N \times (QS)}$. 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 qth angle sample is $\Theta_q = (2q - 1 - Q)/Q$ for $q = 1, 2, \ldots, Q$. The sth distance sample is

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

for $s = 1, 2, \ldots, S$, where β is a factor to adjust the coherence of the adjacent codewords [13]. By setting $d_{(Q+1)/2,S} = 0.5\sqrt{D^3/\lambda}$, we have $\beta = \sqrt{\frac{N}{2S^2}}$. Then (11) is rewritten as

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

Then we let $\boldsymbol{C}_{\mathrm{n}} \triangleq \{\boldsymbol{C}_1, \boldsymbol{C}_2, \ldots, \boldsymbol{C}_Q\}$, where the sth column of $[\boldsymbol{C}_q]$ is $[\boldsymbol{C}_q]_{:,s} = \boldsymbol{\alpha}(N, \Theta_q, d_{q,s})$. Accordingly, the hybrid-field codebook is defined as

$$\boldsymbol{C}_{\mathrm{h}} \triangleq \{\boldsymbol{C}_{\mathrm{n}}, \boldsymbol{C}_{\mathrm{f}}\} \in \mathbb{C}^{N \times (QS+Q)}. \tag{13}$$

Based on (1), we have

$$y_p = [\boldsymbol{C}_{\mathrm{h}}]_{:,p}^{\mathrm{H}} \boldsymbol{h} x_p + [\boldsymbol{C}_{\mathrm{h}}]_{:,p}^{\mathrm{H}} \boldsymbol{\eta} \tag{14}$$

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

$$\widetilde{p} = \arg \max_{p=1,2,\ldots,QS+Q} \left| [\boldsymbol{C}_{\mathrm{h}}]_{:,p}^{\mathrm{H}} \boldsymbol{h} \right|. \tag{15}$$

To solve (15), we need to test all the codewords in $\boldsymbol{C}_{\mathrm{h}}$ one by one, which is called the hybrid-field beam sweeping. To perform the hybrid-field beam sweeping, $(QS + Q)$ 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.

We define an analog combiner as

$$\boldsymbol{W} \triangleq \text{blkdiag}\{\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_{N_{\text{RF}}}\} \tag{16}$$

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

$$\max_{\boldsymbol{W}, \boldsymbol{v}} \; \left| \boldsymbol{v} \boldsymbol{W} \boldsymbol{h} \right|$$

$$\text{s.t.} \quad \|\boldsymbol{v}\boldsymbol{W}\|_2 = 1, \; \left| [\boldsymbol{w}_t]_m \right| = 1 \tag{17}$$

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

$$\max_{\boldsymbol{W}, \boldsymbol{v}} \; \left| \boldsymbol{v} \boldsymbol{W} \boldsymbol{\alpha}(N, \Omega, r) \right| \tag{18a}$$

$$\text{s.t.} \quad \|\boldsymbol{v}\boldsymbol{W}\|_2 = 1, \; \left| [\boldsymbol{w}_t]_m \right| = 1 \tag{18b}$$

for $m = 1, 2, \ldots, M$ and $t = 1, 2, \ldots, N_{\text{RF}}$. In (18a), the subscript "l" is omitted for the simplification for the rest of this paper.

According to the Cauchy-Schwartz inequality, we have

$$\left| \boldsymbol{v} \boldsymbol{W} \boldsymbol{\alpha}(N, \Omega, r) \right| \leq \|\boldsymbol{v}\|_2 \; \|\boldsymbol{W} \boldsymbol{\alpha}(N, \Omega, r)\|_2. \tag{19}$$

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

The design of $\boldsymbol{W}$ can be formulated as

$$\max_{\boldsymbol{W}} \; \left\| \boldsymbol{W} \boldsymbol{\alpha}(N, \Omega, r) \right\|_2$$

$$\text{s.t.} \quad \left| [\boldsymbol{w}_t]_m \right| = 1 \tag{20}$$

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

$$\max_{\boldsymbol{w}_t} \ \left| \boldsymbol{w}_t \boldsymbol{G}_t \boldsymbol{\alpha}(N, \Omega, r) \right|$$

$$\text{s.t.} \ \ \left| [\boldsymbol{w}_t]_m \right| = 1 \tag{21}$$

for $m = 1, 2, \ldots, M$, where we define $\boldsymbol{G}_t \triangleq [\boldsymbol{0}_{M \times (t-1)M}, \boldsymbol{I}_M, \boldsymbol{0}_{M \times (N_{\mathrm{RF}}-t)M}]$. The optimal solution of (21) is

$$\overline{\boldsymbol{w}}_t = \sqrt{N} \big(\boldsymbol{G}_t \boldsymbol{\alpha}(N, \Omega, r) \big)^{\mathrm{H}} \tag{22}$$

for $t = 1, 2, \ldots, N_{\mathrm{RF}}$.

Note that $\boldsymbol{G}_t \boldsymbol{\alpha}(N, \Omega, r)$ is essentially the channel steering vector between the user and the tth subarray. From (2), the whole 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 $1/N_{\mathrm{RF}}^2$ of the Rayleigh distance of the whole array at the BS, which implies that a user in the near field of the whole array might be in the far field of a subarray. For example, if $N = 256$, $\lambda = 0.003$ m, and $N_{\mathrm{RF}} = 4$, the Rayleigh distance of the whole 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 whole array. Based on the above discussion, we aim 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

$$\Gamma_{\mathrm{max}} = \max\{1 - N_{\mathrm{RF}}/\sqrt[4]{2N}, 0\}. \tag{23}$$

Lemma 2: Denote the sine result of the angle that points from the center of the tth subarray to the user as Ψ_t and denote the beam center of $\boldsymbol{G}_t \boldsymbol{\alpha}(N, \Omega, r)$ 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 $N_{\mathrm{RF}} = 4$,

the maximum beam gain loss is only 16%. From **Lemma 2**, considerable beamforming gain can be obtained if the tth subarray receives the signal from channel $\boldsymbol{G}_t \boldsymbol{\alpha}(N, \Omega, r)$ with $\boldsymbol{\beta}(M, \Psi_t)$.

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

$$\widehat{\boldsymbol{w}}_t = \sqrt{M}\boldsymbol{\beta}(M, \Psi_t)^{\mathrm{H}} \tag{24}$$

for $t = 1, 2, \ldots, N_{\mathrm{RF}}$ to approximate the near-field channel steering vectors in (22). Therefore, we can use (24) for both the far-field and near-field channels. By substituting (24) into (16), we express the designed analog combiner for (18) as

$$\widehat{\boldsymbol{W}} = \mathrm{blkdiag}\{\widehat{\boldsymbol{w}}_1, \widehat{\boldsymbol{w}}_2, \ldots, \widehat{\boldsymbol{w}}_{N_{\mathrm{RF}}}\}. \tag{25}$$

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

$$\max_{\boldsymbol{v}} \ \boldsymbol{v}\widehat{\boldsymbol{W}}\boldsymbol{\alpha}(N, \Omega, r) \tag{26a}$$

$$\text{s.t.} \ \ \|\boldsymbol{v}\|_2 = 1/\sqrt{N}. \tag{26b}$$

Note that (18a) can be rewritten as (26a) because we can always adjust the phase of $\boldsymbol{v}$ so that $\boldsymbol{v}\widehat{\boldsymbol{W}}\boldsymbol{\alpha}(N, \Omega, r)$ is real positive and the maximum of $|\boldsymbol{v}\widehat{\boldsymbol{W}}\boldsymbol{\alpha}(N, \Omega, r)|$ is still achieved. The optimal $\boldsymbol{v}$ for (26) is

$$\widehat{\boldsymbol{v}} = \frac{\big(\widehat{\boldsymbol{W}}\boldsymbol{\alpha}(N, \Omega, r)\big)^{\mathrm{H}}}{\sqrt{N}\big\|\widehat{\boldsymbol{W}}\boldsymbol{\alpha}(N, \Omega, r)\big\|_2}. \tag{27}$$

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

$$\boldsymbol{\Phi} = \{\sqrt{M}\boldsymbol{\beta}(M, \Phi_1), \sqrt{M}\boldsymbol{\beta}(M, \Phi_2), \ldots, \sqrt{M}\boldsymbol{\beta}(M, \Phi_M)\}, \tag{28}$$

where

$$\Phi_m = (2m - 1 - M)/M \tag{29}$$

for $m = 1, 2, \ldots, M$. In fact, the DFT codebook equally samples the full space $[-1, 1]$ by M angles, where the mth angle is Φ_m. For the tth subarray, the index of the codeword in $\boldsymbol{\Phi}$ best fitting for the channel is

$$\widetilde{m}_t = \arg \min_{m=1,2,\ldots,M} |\Phi_m - \Psi_t|. \tag{30}$$

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

$$\widetilde{\boldsymbol{W}} = \mathrm{blkdiag}\{\widetilde{\boldsymbol{w}}_1, \widetilde{\boldsymbol{w}}_2, \ldots, \widetilde{\boldsymbol{w}}_{N_{\mathrm{RF}}}\}, \tag{31}$$

$$\widetilde{\boldsymbol{v}} = \frac{\left(\widetilde{\boldsymbol{W}}\boldsymbol{\alpha}(N, \Omega, r)\right)^{\mathrm{H}}}{\sqrt{N}\left\|\widetilde{\boldsymbol{W}}\boldsymbol{\alpha}(N, \Omega, r)\right\|_2}, \tag{32}$$

where

$$\widetilde{\boldsymbol{w}}_t = [\boldsymbol{\Phi}]^{\mathrm{H}}_{:,\widetilde{m}_t} \tag{33}$$

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

Note that for the partially-connected structure, each subarray is solely connected to an RF chain, which indicates that each RF chain can support an independent beam training based on a subarray. As a result, we need totally M times of beam training. In particular, 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.

Now we propose a two-stage hybrid-field beam training scheme. In the first stage, each subarray independently uses M far-field channel steering vectors for analog combining, where the mth combiner according to (28) is

$$\overline{\boldsymbol{W}}_m = \mathrm{blkdiag}\{\sqrt{M}\boldsymbol{\beta}(M, \Phi_m)^{\mathrm{H}}, \ldots, \sqrt{M}\boldsymbol{\beta}(M, \Phi_m)^{\mathrm{H}}\}. \tag{34}$$

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

$$\boldsymbol{z}_m = \overline{\boldsymbol{W}}_m \boldsymbol{h} x_m + \overline{\boldsymbol{W}}_m \boldsymbol{\eta} \tag{35}$$

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

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

$$[\boldsymbol{C}_{\mathrm{h}}]_{:,p} = \begin{cases} \boldsymbol{\alpha}\left(N, \Theta_{\bar{q}}, d_{\bar{q},\bar{s}}\right), & p \le QS, \\ \boldsymbol{\beta}\left(N, \frac{2(p-QS)-1-Q}{Q}\right), & p > QS, \end{cases} \tag{36}$$

where

$$\bar{q} = \lceil p/S \rceil \text{ and } \bar{s} = p - (\bar{q} - 1)S. \tag{37}$$

Suppose $[\boldsymbol{C}_{\mathrm{h}}]_{:,p}$ is the codeword best fitting for the channel. We will design analog combiner $\boldsymbol{F}_p$ and digital combiner $\widetilde{\boldsymbol{v}}_p$ so that the phase differences of subarrays induced by the near-field effect can be compensated and the beam gain of $\widetilde{\boldsymbol{v}}_p \boldsymbol{F}_p$ can approach that of $[\boldsymbol{C}_{\mathrm{h}}]_{:,p}$. Replacing $\boldsymbol{\alpha}(N, \Omega, r)$ in (18) by $[\boldsymbol{C}_{\mathrm{h}}]_{:,p}$, we can obtain $\widetilde{\boldsymbol{w}}_t^{(p)}$ via (33). Similar to (31) and (32), we design the analog combiner and the digital combiner, respectively, as

$$\boldsymbol{F}_p = \mathrm{blkdiag}\{\widetilde{\boldsymbol{w}}_1^{(p)}, \widetilde{\boldsymbol{w}}_2^{(p)}, \ldots, \widetilde{\boldsymbol{w}}_{N_{\mathrm{RF}}}^{(p)}\}, \tag{38}$$

$$\widetilde{\boldsymbol{v}}_p = \frac{(\boldsymbol{F}_p [\boldsymbol{C}_{\mathrm{h}}]_{:,p})^{\mathrm{H}}}{\sqrt{N}\big\|\boldsymbol{F}_p [\boldsymbol{C}_{\mathrm{h}}]_{:,p}\big\|_2}. \tag{39}$$

Note that both $\boldsymbol{F}_p$ and $\widetilde{\boldsymbol{v}}_p$ can be computed offline before the beam training, which can substantially reduce the computational complexity of the beam training. If we use $\boldsymbol{F}_p$ and $\widetilde{\boldsymbol{v}}_p$ for combining, we have

$$\widetilde{\boldsymbol{z}}_p = \boldsymbol{F}_p \boldsymbol{h} x_p + \boldsymbol{F}_p \boldsymbol{\eta}, \tag{40}$$

$$\widetilde{y}_p = \widetilde{\boldsymbol{v}}_p \widetilde{\boldsymbol{z}}_p, \ p = 1, 2, \ldots, QS + Q. \tag{41}$$

In fact, $\widetilde{\boldsymbol{z}}_p$ can not be obtained because we do not really perform beam training with $\boldsymbol{F}_p$. However, each entry of $\widetilde{\boldsymbol{z}}_p$ can be obtained from the beam training in (35) because both $\boldsymbol{F}_p$ and $\overline{\boldsymbol{W}}_m$ are composed of channel steering vectors from the same set $\boldsymbol{\Phi}$ in (28). We can obtain $\widetilde{\boldsymbol{z}}_p$ by setting

$$[\widetilde{\boldsymbol{z}}_p]_t = [\boldsymbol{z}_{\widetilde{m}_t}]_t, \tag{42}$$

where $\widetilde{m}_t$ can be obtained from (30) during the design of $\boldsymbol{F}_p$ in (38). Note that (41) can be computed in parallel to speed up the beam training. From $\{\widetilde{y}_1, \widetilde{y}_2, \ldots, \widetilde{y}_{QS+Q}\}$, we select one with the largest power, which can be expressed as

$$\widetilde{p} = \arg \max_{p=1,2,\ldots,QS+Q} \big|\widetilde{y}_p\big|^2. \tag{43}$$

From the hybrid-field codebook, $\boldsymbol{C}_{\mathrm{h}}$, we select the codeword $[\boldsymbol{C}_{\mathrm{h}}]_{:,\widetilde{p}}$ corresponding to the dedicated digital combiner $\widetilde{\boldsymbol{v}}_{\widetilde{p}}$ that can achieve the largest combining power $|\widetilde{y}_{\widetilde{p}}|^2$ as the result of the THBT. Moreover, if the user is in the near field, we can also roughly locate the user

---

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

---

1: **Input:** N, N_{RF}, M, S, λ.

2: **First Stage:**

3: Obtain $\boldsymbol{z}_m$, $m = 1, 2, \ldots, M$ via (35).

4: **Second Stage:**

5: **for** $p = 1, 2, \ldots, QS + Q$ **do**

6:     Obtain $[\boldsymbol{C}_{\mathrm{h}}]_{:,p}$ via (36).

7:     Obtain $\boldsymbol{F}_p$ and $\widetilde{\boldsymbol{v}}_p$ via (38) and (39), respectively.

8:     Obtain $\widetilde{\boldsymbol{z}}_p$ via (42).

9:     Obtain $\widetilde{y}_p$ via (41).

10: **end for**

11: Obtain $\widetilde{p}$ via (43).

12: Obtain $\widetilde{\Omega}$ and $\widetilde{r}$ via (45).

13: **Output:** $[\boldsymbol{C}_{\mathrm{h}}]_{:,\widetilde{p}}$, $\widetilde{\Omega}$ and $\widetilde{r}$.

---

based on the beam training results. According to (37), the indices of the quantized angle and the quantized distance can be expressed as

$$\widetilde{q} = \lceil \widetilde{p}/S \rceil \text{ and } \widetilde{s} = \widetilde{p} - (\widetilde{q} - 1)S. \tag{44}$$

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

$$\widetilde{\Omega} = (2\widetilde{q} - 1 - Q)/Q \text{ and } \widetilde{r} = \frac{N^{3/2}\lambda S(1 - \Theta_{\widetilde{q}}^2)}{4\sqrt{2\widetilde{s}}}. \tag{45}$$

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

Compared to the hybrid-field beam sweeping based on $\boldsymbol{C}_{\mathrm{h}}$ in (13) that needs $QS + Q$ times of beam training, the training overhead of the THBT scheme is substantially reduced to M. The computational complexity of the proposed scheme mainly comes from step 9 in **Algorithm 1** and is totally $\mathcal{O}(N_{\mathrm{RF}}Q(S + 1))$.

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

$$\widetilde{k} = -\frac{\lambda(1 - \widetilde{\Omega}^2)}{2\widetilde{r}} \text{ and } \widetilde{b} = \widetilde{\Omega} - \widetilde{k}(N+1). \tag{46}$$

Note that the resolutions of $\widetilde{k}$ and $\widetilde{b}$ 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

$$\overline{\overline{W}} = \text{blkdiag}\{\overline{\overline{w}}_1^{\text{H}}, \overline{\overline{w}}_2^{\text{H}}, \dots, \overline{\overline{w}}_{N_{\text{RF}}}^{\text{H}}\}, \tag{47}$$

where

$$[\overline{\overline{w}}_t]_m = e^{j\pi(\widetilde{k}((t-1)M+m)^2 + \widetilde{b}((t-1)M+m))} \tag{48}$$

for $m = 1, 2, \dots, M$ and $t = 1, 2, \dots, N_{\text{RF}}$, so that the near-field effect can be alleviated. Note that $\overline{\overline{W}}$ is designed based on $\widetilde{k}$ and $\widetilde{b}$. Then the received signal of the tth subarray after analog combining can be expressed as

$$
\begin{aligned}
\widehat{z}_t &\overset{\text{(a)}}{\approx} \sum_{n=(t-1)M+1}^{tM} e^{-j\pi(\widetilde{k}n^2 + \widetilde{b}n)}[\boldsymbol{h}]_n \\
&\overset{\text{(b)}}{\approx} g \sum_{n=(t-1)M+1}^{tM} e^{-j\pi(\widetilde{k}n^2 + \widetilde{b}n)}[\boldsymbol{\alpha}(N, \Omega, r)]_n \\
&\overset{\text{(c)}}{\approx} \widetilde{g} \sum_{n=(t-1)M+1}^{tM} e^{-j\pi(\widetilde{k}n^2 + \widetilde{b}n)}[\boldsymbol{\gamma}(N, \Omega, r)]_n \\
&= \widetilde{g} \sum_{n=(t-1)M+1}^{tM} e^{j\pi(\Delta k n^2 + \Delta b n)} \\
&= \widetilde{g} e^{j\pi(\Delta k (t-1)^2 M^2 + \Delta b(t-1)M)} \cdot \sum_{m=1}^{M} e^{j\pi(\Delta k m^2 + \Delta b m + 2\Delta k m M(t-1))},
\end{aligned} \tag{49}
$$

for $t = 1, 2, \dots, N_{\text{RF}}$, where $\Delta k \triangleq k - \widetilde{k}$, $\Delta b \triangleq b - \widetilde{b}$ and $\widetilde{g} \triangleq g e^{j2\pi(\Omega(\delta_1 - 1/2) - \rho(N+1)^2/16)}$. In (49), 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 (49) is tricky due to the quadratic phase term $\Delta k m^2$. To obtain a deeper insight of $\widehat{z}_t$, we first

$$\widehat{z}_t \approx \widetilde{g}e^{j\pi(\Delta k(t-1)^2 M^2 + \Delta b(t-1)M)} \cdot \sum_{m=1}^{M} \left(\frac{1}{2} \int_{2\Delta kM}^{2\Delta k} e^{-j\pi(\frac{\omega^2}{4\Delta k} + \frac{1}{4})} \sqrt{\frac{1}{-\Delta k}} e^{j\pi m\omega} \mathrm{d}\omega \right) e^{j\pi(\Delta bm + 2\Delta kmM(t-1))}$$

$$= \widetilde{g}e^{j\pi(\Delta k(t-1)^2 M^2 + \Delta b(t-1)M)} \frac{1}{2} \int_{2\Delta kM}^{2\Delta k} e^{-j\pi(\frac{\omega^2}{4\Delta k} + \frac{1}{4})} \sqrt{\frac{1}{-\Delta k}} \left(\sum_{m=1}^{M} e^{j\pi(\Delta bm + 2\Delta kmM(t-1))} e^{j\pi m w} \right) \mathrm{d}\omega$$

$$= \overline{g}e^{j\pi(\Delta k(t-1)^2 M^2 + \Delta b(t-1)M)} \int_{2\Delta kM}^{2\Delta k} e^{-j\pi(\frac{\omega^2}{4\Delta k} + \frac{1}{4})} \left(e^{j(M+1)(\pi\phi_t + \pi\omega)/2} \frac{\sin(M(\pi\phi_t + \pi\omega)/2)}{\sin((\pi\phi_t + \pi\omega)/2)} \right) \mathrm{d}\omega$$

$$= \overline{g}e^{j\pi(\Delta \widetilde{k}(t-1)^2 + \Delta \widetilde{b}(t-1))} \int_{2\Delta kM}^{2\Delta k} \underbrace{e^{j\pi\left(\frac{(M+1)\omega}{2} - \frac{\omega^2}{4\Delta k}\right)}}_{\mathcal{P}(\omega)} \underbrace{\frac{\sin(M(\pi\phi_t + \pi\omega)/2)}{\sin((\pi\phi_t + \pi\omega)/2)}}_{\mathcal{A}(\phi_t, \omega)} \mathrm{d}\omega$$

$$= \overline{g}\underbrace{e^{j\pi(\Delta \widetilde{k}(t-1)^2 + \Delta \widetilde{b}(t-1))}}_{\mathcal{C}(t)} \underbrace{\int_{2\Delta kM}^{2\Delta k} \mathcal{P}(\omega)\mathcal{A}(\phi_t, \omega)\mathrm{d}\omega}_{\mathcal{B}(\phi_t)} \tag{52}$$

focus on $e^{j\pi\Delta km^2}$. Define $\boldsymbol{s} \triangleq [e^{j\pi\Delta k}, e^{j\pi4\Delta k}, \ldots, e^{j\pi M^2\Delta k}]^{\mathrm{T}}$. According to the PSP [30]–[32], the discrete-time Fourier transform of $\boldsymbol{s}$ is expressed as

$$\mathcal{F}(\omega) = \sum_{m=1}^{M}[\boldsymbol{s}]_m e^{-j\pi m\omega} \approx \begin{cases} e^{-j\pi(\frac{\omega^2}{4\Delta k} + \frac{1}{4})}\sqrt{\frac{1}{-\Delta k}}, & 2\Delta kM \leq \omega \leq 2\Delta k, \\ 0, & \text{others}, \end{cases} \tag{50}$$

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

$$[\boldsymbol{s}]_m = e^{j\pi\Delta km^2} \approx \frac{1}{2}\int_{-1}^{1} \mathcal{F}(\omega)e^{j\pi m\omega}\mathrm{d}\omega = \frac{1}{2}\int_{2\Delta kM}^{2\Delta k} e^{-j\pi(\frac{\omega^2}{4\Delta k} + \frac{1}{4})}\sqrt{\frac{1}{-\Delta k}}e^{j\pi m\omega}\mathrm{d}\omega. \tag{51}$$

Substituting (51) into (49), we have (52), which is shown at the top of this page, where

$$\overline{g} \triangleq \frac{1}{2\sqrt{-\Delta k}}\widetilde{g}e^{j\pi((M+1)\Delta b/2 - 1/4)}, \ \Delta\widetilde{k} \triangleq \Delta kM^2,$$

$$\Delta\widetilde{b} \triangleq (\Delta b + \Delta k(M+1))M, \ \text{and} \ \phi_t \triangleq \Delta b + 2\Delta kM(t-1). \tag{53}$$

Remark 1: From (52), the power of $\widehat{z}_t$ 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 w will not cancel

each other. Then we have

$$|\phi_t + \omega| \le \frac{1}{M}, \tag{54a}$$

$$\left| \frac{(M+1)\omega}{2} - \frac{\omega^2}{4\Delta k} \right| \le 1/2, \tag{54b}$$

for $\omega \in [2\Delta k M, 2\Delta k]$. Substituting ϕ_t in (53) into (54a), we have

$$|\Delta k| \le \frac{1}{M(N-1)} - \frac{1}{Q(N-1)}. \tag{55}$$

From (54b), we have

$$|\Delta k| \le \frac{1}{4M^2}. \tag{56}$$

Note that

$$\frac{1}{M(N-1)} - \frac{1}{Q(N-1)} < \frac{1}{M(N-1)} \approx \frac{1}{N_{\mathrm{RF}}M^2} \le \frac{1}{4M^2} \tag{57}$$

for $N_{\mathrm{RF}} \ge 4$, which is easy to be satisfied for XL-MIMO systems. Therefore, we omit the constraint in (56) and focus on (55) to simplify the expression. Note that

$$|\Delta k| \approx \frac{1}{4} \left| \frac{\lambda(1 - \Theta_q^2)}{2d_{q,s}} - \frac{\lambda(1 - \Theta_q^2)}{2d_{q,s+1}} \right| = \frac{\sqrt{2}}{2N^{3/2}S}. \tag{58}$$

Combining (56) and (58), we have

$$Q \ge M \quad \text{and} \quad S \ge \frac{\sqrt{2}(N-1)}{2N^{3/2}\left(\frac{1}{M} - \frac{1}{Q}\right)}, \tag{59}$$

which provides guidance for the settings of S and Q.

In (52), the expression of $\widehat{z}_t$ is analytically derived, where the phase of $\widehat{z}_t$ shifts along with t. Specifically, the phase of $\widehat{z}_t$ is related to $\overline{g}$, $\mathcal{C}(t)$, and $\mathcal{B}(\phi_t)$, where $\overline{g}$ 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 $\mathcal{A}(\phi_t, \omega)$ is always positive under the constraint in (54a). Therefore, the unwrapped phase of $\widehat{z}_t$ 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 $\Upsilon_t \triangleq \angle \widehat{z}_t$, where $\angle(\cdot)$ denotes the phase of a complex value. The first-order difference of Υ_t can be expressed as

$$\Delta \Upsilon_t \triangleq \Upsilon_{t+1} - \Upsilon_t = \pi(\Delta \widetilde{k}(2t-1) + \Delta \widetilde{b}), \tag{60}$$

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

$$\Delta^2 \Upsilon_t \triangleq \Delta \Upsilon_{t+1} - \Delta \Upsilon_t = 2\pi \Delta \widetilde{k}, \tag{61}$$

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

$$|\Delta^2 \Upsilon_t| = 2\pi |\Delta \widetilde{k}| \overset{(a)}{<} \frac{2\pi M^2}{M(N-1)} = \frac{2\pi}{N_{\mathrm{RF}}} \cdot \frac{N}{N-1} \overset{(b)}{<} \pi \tag{62}$$

where (a) can be obtained via (55) and (b) holds for $N_{\mathrm{RF}} > 2$. To avoid phase wrap, we can shift $\Delta^2 \Upsilon_t$ to $[-\pi, \pi]$ by

$$\overline{\Upsilon}_t \triangleq \mathrm{mod}(\Delta^2 \Upsilon_t + \pi, 2\pi) - \pi. \tag{63}$$

Then the estimation of Δk can be expressed as

$$\Delta \widehat{k} = \frac{1}{N_{\mathrm{RF}} - 2} \sum_{t=1}^{N_{\mathrm{RF}}-2} \frac{\overline{\Upsilon}_t}{2\pi M^2}. \tag{64}$$

Now we turn to the estimation of Δb. Define

$$\widetilde{\Upsilon}_t \triangleq \Delta \Upsilon_t - (2t-1)M^2 \Delta \widehat{k} \pi - M(M+1)\Delta \widehat{k} \pi. \tag{65}$$

Then according to (60), we have

$$\widetilde{\Upsilon}_t = M \Delta b \pi + 2u\pi, \ u \in \mathbb{N}. \tag{66}$$

Note that

$$|M \Delta b \pi| \leq \frac{M\pi}{Q} + \frac{(N+1)\pi}{(N-1)} - \frac{M(N+1)\pi}{Q(N-1)} \overset{(a)}{\approx} \pi. \tag{67}$$

where (a) holds because $(N+1)/(N-1) \approx 1$ for XL-MIMO systems. Therefore, we can also shift the results in (66) to $[-\pi, \pi]$ to avoid phase wrap. Define $\widehat{\Upsilon}_t \triangleq \mathrm{mod}(\widetilde{\Upsilon}_t + \pi, 2\pi) - \pi$. Then the estimation of Δb can be expressed as

$$\Delta \widehat{b} = \frac{1}{N_{\mathrm{RF}} - 1} \sum_{t=1}^{N_{\mathrm{RF}}-1} \frac{\widehat{\Upsilon}_t}{M\pi}. \tag{68}$$

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

$$\widehat{k} = \Delta \widehat{k} + \widetilde{k} \ \text{and} \ \widehat{b} = \Delta \widehat{b} + \widetilde{b}. \tag{69}$$

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

$$\widehat{\Omega} = \widehat{b} + \widehat{k}(N+1) \ \text{and} \ \widehat{r} = -\frac{\lambda(1 - \widehat{\Omega}^2)}{2\widehat{k}}. \tag{70}$$

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

---

1: **Input:** N, N_{RF}, M, S, λ, $\overline{z}_t$, $\widetilde{\Omega}$ and $\widetilde{r}$.

2: Obtain $\widetilde{k}$ and $\widetilde{b}$ via (46).

3: Obtain $\Upsilon_t = \angle \overline{z}_t$, for $t = 1, 2, \ldots, N_{\mathrm{RF}}$.

4: Obtain $\Delta \Upsilon_t$ and $\Delta^2 \Upsilon_t$ via (60) and (61), respectively.

5: Obtain $\Delta \widehat{k}$ and $\Delta \widehat{b}$ via (64) and (68), respectively.

6: Obtain $\widehat{k}$ and $\widehat{b}$ via (69).

7: Obtain $\widehat{\Omega}$ and $\widehat{r}$ via (70).

8: **Output:** $\widehat{k}$, $\widehat{b}$, $\widehat{\Omega}$ and $\widehat{r}$.

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

Remark 2: The overheads, computation 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.
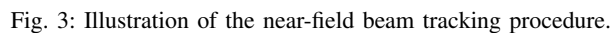
The proposed BRPSS has closed-form expressions, which no longer needs running any algorithms and therefore has very low computation complexity. The computation of the BRPSS mainly comes from the computing of (60), (61), (64) and (68), which are all only related to N_{RF}. Therefore, the computation complexity of the BRPSS is $\mathcal{O}(N_{\mathrm{RF}})$.

Note that the unwrapped phase of $\widehat{z}_t$ is the quadratic function of the indices of subarrays. Therefore, at least 3 subarrays are needed for the BRPSS scheme.

Remark 3: The BRPSS is developed based on the phase shifts of the subarrays. Therefore, it can also be extended to the fully-connected structure if the overlapped or non-overlapped subarrays can be constructed by deactivating part of antennas.

V. NEAR-FIELD BEAM TRACKING

In this section, we develop a low-complexity near-field beam tracking (NFBT) scheme, where a kinematic model is adopted to characterize the channel variations at different time slots and the BRPSS scheme is used to estimate the real-time channel parameters. Then the kinematic model

Fig. 3: Illustration of the near-field beam tracking procedure.

and real-time estimates are exploited by the EKF to track and predict the near-field channel parameters.

Note that the near-field channel is related to both the angle and the distance of the user, which facilitates the implementation of beam tracking. Specifically, based on the estimated angle and distance of the user, the position of the user can be calculated. Therefore, the kinematic model can be exploited to characterize the channel variations at different time slots. In Fig. 3, 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. Denote the kinematic parameters of the user at the ith block as $\boldsymbol{p}[i] = [a_x[i], a_y[i], v_x[i], v_y[i]]^{\mathrm{T}}$, 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 ith 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 initialize kinematic parameters of the user as $\boldsymbol{p}[0] = [a_x[0], a_y[0], 0, 0]^{\mathrm{T}}$, which can be obtained via the THBT and BRPSS schemes.

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

---

Algorithm 3 Near-Field Beam Tracking (NFBT) Scheme

1: **Input:** N, N_{RF}, M, S, λ, $\boldsymbol{p}[0]$ and ΔT.

2: **Initialization:** $i \leftarrow 1$.

3: **while** stop conditions in Sec. V-5 are not satisfied **do**

4:      Obtain $\overline{\boldsymbol{p}}[i]$ via (71).

5:      Obtain $\widehat{\boldsymbol{p}}[i]$ via (77).

6:      Obtain $\widetilde{\boldsymbol{p}}[i]$ via the EKF.

7:      Obtain $\widetilde{\boldsymbol{h}}_i$ via (78).

8:      $i \leftarrow i + 1$.

9: **end while**

10: **Output:** $\widetilde{\boldsymbol{h}}_i$.

---

expressed as

$$\overline{\boldsymbol{p}}[i] = \underbrace{\begin{bmatrix} 1 & 0 & \Delta T & 0 \\ 0 & 1 & 0 & \Delta T \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}}_{\boldsymbol{\Xi}} \widetilde{\boldsymbol{p}}[i-1] \tag{71}$$

where the predicted distance and angle can be expressed as

$$\overline{\zeta}_i = \sqrt{\overline{a}_x[i]^2 + \overline{a}_y[i]^2} \text{ and } \overline{\vartheta}_i = \arctan(\overline{a}_y[i]/\overline{a}_x[i]). \tag{72}$$

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

$$\overrightarrow{\boldsymbol{W}}_i = \mathrm{blkdiag}\{\overrightarrow{\boldsymbol{w}}_1^{\mathrm{H}}, \overrightarrow{\boldsymbol{w}}_2^{\mathrm{H}}, \ldots, \overrightarrow{\boldsymbol{w}}_{N_{\mathrm{RF}}}^{\mathrm{H}}\}, \tag{73}$$

where

$$[\overrightarrow{\boldsymbol{w}}_t]_m = e^{j\pi(\overline{k}_i((t-1)M+m)^2 + \overline{b}_i((t-1)M+m))} \tag{74}$$

for $m = 1, \ldots, M$ and $t = 1, 2, \ldots, N_{\mathrm{RF}}$. In (74),

$$\overline{k} = -\frac{\lambda(1 - \sin^2 \overline{\vartheta}_i)}{2\overline{\zeta}_i} \text{ and } \overline{b} = -\sin \overline{\vartheta}_i - \overline{k}(N+1) \tag{75}$$

which are calculated based on the predicted distance and angle in (72). Then the received signal of the tth subarray after analog combining can be expressed as

$$\overline{z}_t = \overrightarrow{\boldsymbol{w}}_t^{\mathrm{H}} \boldsymbol{G}_t \overline{\boldsymbol{h}}_i + \overrightarrow{\boldsymbol{w}}_t^{\mathrm{H}} \boldsymbol{G}_t \overline{\boldsymbol{\eta}}_i, \tag{76}$$

for $t = 1, 2, \ldots, N_{\mathrm{RF}}$, where $\overline{\boldsymbol{h}}_i$ denotes the channel between the user and the BS at the ith block and $\overline{\boldsymbol{\eta}}_i$ denotes the AWGN. Substituting N, N_{RF}, M, S, λ, $\overline{z}_t$, $-\sin\overline{\vartheta}_i$ and $\overline{\zeta}_i$ into **Algorithm** 2, we can obtain estimated angle $\widehat{\vartheta}_i$ and estimated distance $\widehat{\zeta}_i$ of the user at the ith block. Then the estimated state vector can be expressed as

$$\widehat{\boldsymbol{p}}[i] = \left[\widehat{\zeta}_i \cos\widehat{\vartheta}_i, \widehat{\zeta}_i \sin\widehat{\vartheta}_i, 0, 0\right]^{\mathrm{T}}. \tag{77}$$

4) Filtering: The estimated stated vector, $\widehat{\boldsymbol{p}}[i]$, and the predicted state vector, $\overline{\boldsymbol{p}}[i]$, can be exploited by the widely-adopted EKF for effective tracking. Since the EKF has been exhaustively introduced in a variety of works [33]–[37], we omit the details and denote the filtered state vector at the ith block after performing the EKF as $\widetilde{\boldsymbol{p}}[i]$. Then the estimated LoS channel at the ith block can be expressed as

$$[\widetilde{\boldsymbol{h}}_i]_n = e^{j\pi\left(\overrightarrow{k} n^2 + \overrightarrow{b} n\right)}, \quad \overrightarrow{k} = -\frac{\lambda(1 - \sin^2\widetilde{\vartheta}_i)}{2\widetilde{\zeta}_i}, \quad \overrightarrow{b} = -\sin\widetilde{\vartheta}_i - \overrightarrow{k}(N + 1),$$

$$\widetilde{\zeta}_i = \sqrt{\widetilde{a}_x[i]^2 + \widetilde{a}_y[i]^2}, \text{ and } \widetilde{\vartheta}_i = \arctan(\widetilde{a}_y[i]/\widetilde{a}_x[i]), \tag{78}$$

for $n = 1, 2, \ldots, N$, which can be used to design the hybrid precoders for the 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.

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

VI. SIMULATION RESULTS

Now we evaluate the performance of the proposed schemes. We consider an XL-MIMO system equipped with $N = 512$ antennas. The antenna array is composed of $N_{\mathrm{RF}} = 4$ subarrays with each subarray having $M = 128$ antennas. The wavelength is set to be $\lambda = 0.003$ m corresponding to the carrier frequency of 100 GHz. 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 $g_3 \sim \mathcal{CN}(0, 0.01)$. Channel angle Ω_l of the lth path obeys the uniform distribution between $[-\sqrt{3}/2, \sqrt{3}/2]$. To
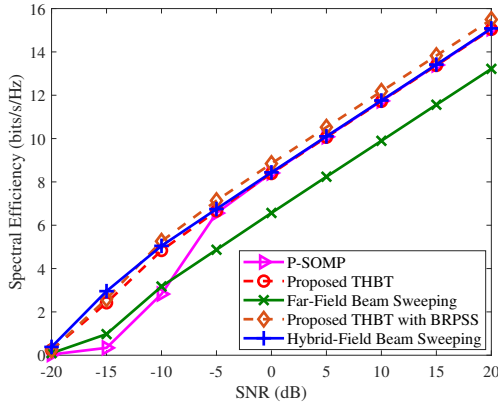
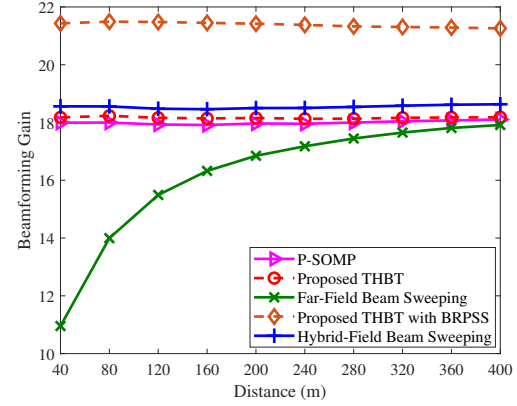Fig. 4: Comparisons of the spectral efficiency for different methods.

Fig. 5: Comparisons of the beamforming gains for different methods.

compare the proposed schemes with P-SOMP [13] fairly, we set the pilot length of P-SOMP to be $M = 128$, which is also the training overhead of the THBT.

A. Evaluation of the THBT and BRPSS Schemes

In Fig. 4, we compare the proposed schemes with P-SOMP, the far-filed beam sweeping [38] and the hybrid-field beam sweeping in terms of spectral efficiency for different SNRs. The distances between the BS and the user or scatterers obey the uniform distribution between $[10, 30]$ m. 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. 4, the hybrid-field beam sweeping can achieve better performance than the other four schemes when the SNR is less than -10 dB, which lies in the fact that hybrid-field beam sweeping exhaustively tests all the codewords in C_h and needs far more times of beam training than the other four 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 four methods 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. Furthermore, the performance of the proposed THBT scheme can approach the performance of the hybrid-field beam sweeping at various SNR conditions. Moreover, the performance of P-SOMP is worse than that of the other four schemes at low SNRs, such as -15 dB, because the random beamforming of P-SOMP cannot achieve enough beamforming gain and will significantly degrade the performance. The performance of the far-field beam sweeping
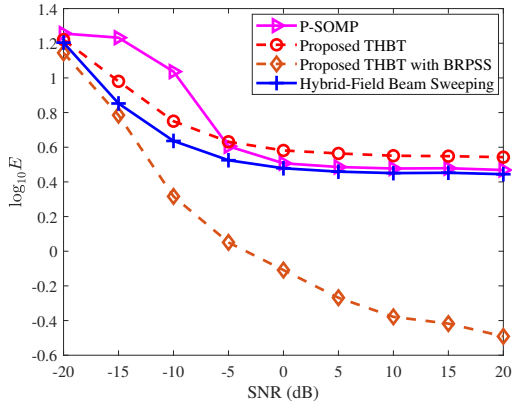
Fig. 6: Comparisons of the positioning error for different methods.
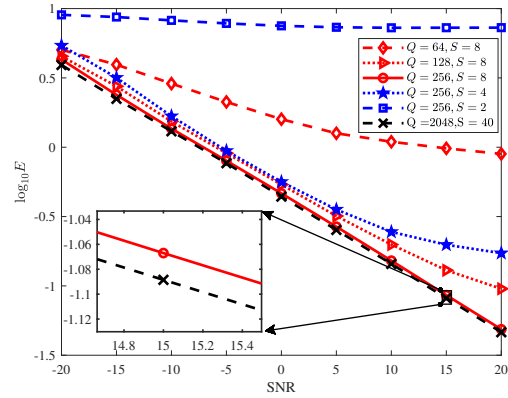

Fig. 7: Evaluation of the BRPSS for different settings of Q and S.

is worse than that of the other four schemes at high SNRs because the beamforming gain of the far-field channel steering vector will decrease in the near field.

In Fig. 5, we compare the proposed schemes with P-SOMP, the far-filed beam sweeping and the hybrid-field beam sweeping in terms of beamforming gain for different distances. The distances between the BS and the user or scatterers obey the uniform distribution between $[10, r]$ m, where r ranges from 40 to 400. The SNR is fixed to be -10 dB. From Fig. 5, 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 far-field beam sweeping will suffer severe loss of beamforming gain because it only considers the far-field channels. By contrast, the other four schemes are robust to the distance because they consider both the near-filed channel and the far-field channel. In addition, the performance of the proposed THBT scheme is better than P-SOMP and can approach that of the hybrid-field beam sweeping with only a slight loss of beamforming gain at different distances.

In Fig. 6, we evaluate the positioning performance of the proposed schemes. Since the far-field beam sweeping cannot obtain the position of the user, we only compare the proposed schemes with P-SOMP and the hybrid-filed beam sweeping. We denote the distance between the real position and the estimated position as E. From Fig. 6, 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 proposed THBT performs better than P-SOMP when SNR is less than -5 dB because the random beamforming of P-SOMP cannot achieve enough beamforming gain and significantly degrade the performance. Moreover,
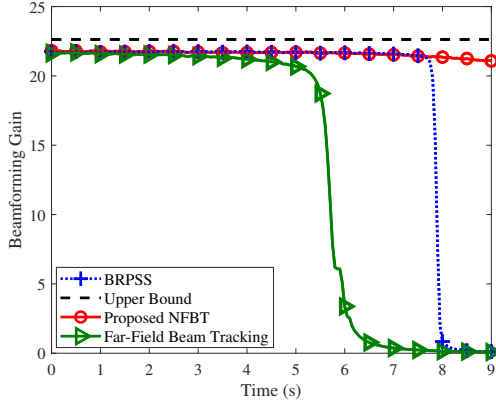
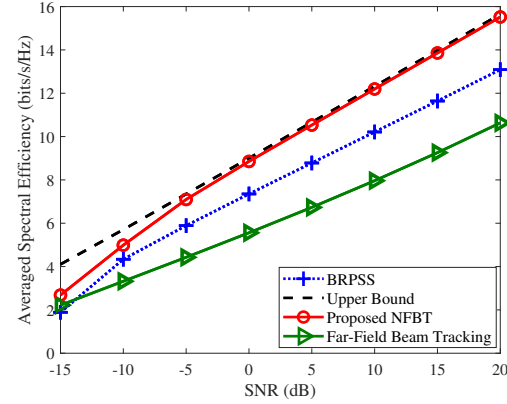Fig. 8: Comparisons of the beamforming gains during the tracking process.

Fig. 9: Comparisons of the averaged spectral efficiency during the tracking process.

P-SOMP performs better than THBT when SNR is larger than -5 dB because P-SOMP exploits the measures of all the pilots and will suffer from high computation complexity while THBT only uses N_{RF} of the measures to form the directional beams.

We also compare the training overheads of different schemes in Fig. 4 to 6. The training overheads of the hybrid-field beam sweeping, the far-field beam sweeping, P-SOMP, the proposed THBT, and the proposed THBT with the BRPSS are $Q(S+1)$, N, M, M, and $M+1$, respectively. Under the simulation settings, these five schemes require 6144, 512, 128, 128, and 129 time slots, respectively, where the proposed THBT can approach the performance of the hybrid-field beam sweeping with 97.92% reduction in training overhead, and the proposed THBT with the BRPSS outperforms the hybrid-field beam sweeping with 97.90% reduction in training overhead.

In Fig. 7, we evaluate the performance of the BRPSS for different settings of Q and S. To remove the effects of the NLoS paths and the prior beam training failure, we set $L = 1$ and suppose the beam training has found the codeword best fitting for the channel in the hybrid-field codebook. From Fig. 7, the positioning error decreases with the increase of S or Q if we fix $Q = 256$ or $S = 8$. This is because beam gain losses of subarrays will occur if the conditions of Q and S in (59) are not satisfied, which will deteriorate the performance of beam refinement. However, if the conditions of Q and S in (59) 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 (59) 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 $\boldsymbol{p}_0 = [50, 50\sqrt{3}, 0, 0]$, $\Delta T = 0.05$ s, $I = 180$ and $[v_x[i], v_y[i]]^{\mathrm{T}} = [-5, -5\sqrt{3}]^{\mathrm{T}}$ for $i = 1, 2, \ldots, I$.

In Fig. 8, we compare the proposed NFBT with the BRPSS and the far-field beam tracking [26] in terms of the beamforming gain during the tracking process, 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 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 far-field beam tracking 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. 9, we compare the proposed NFBT with the BRPSS and the far-field beam tracking in terms of the averaged spectral efficiency during the tracking process for different SNRs. From the figure, the proposed NFBT performs the best among the three schemes and can approach the upper bound when SNR is larger than -5dB, which demonstrates the effectiveness of the proposed schemes.

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 slots

and the EKF has been exploited for beam tracking. Future works will be continued with effective beam training and high-resolution beam refinement for fully-connected structures.

APPENDIX A

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

$$
\begin{aligned}
|G(\boldsymbol{G}_t\boldsymbol{h},\Omega)| &= \sqrt{M}\boldsymbol{\beta}(M,\Omega)^{\mathrm{H}}\boldsymbol{G}_t\boldsymbol{h} \\
&\approx \left| \sum_{n=(t-1)M+1}^{tM} e^{j\pi(kn^2+bn)}e^{-j\pi n\Omega} \right| \\
&= \left| \sum_{n=1}^{M} e^{j\pi(kn^2+\bar{\bar{b}}n)}e^{-j\pi n\Omega} \right| \\
&\approx \left| \int_1^M e^{j\pi(k\chi^2+(\bar{\bar{b}}-\Omega)\chi))}\mathrm{d}\chi \right| \\
&= \left| \int_{-\infty}^{\infty} U(\chi)e^{j\pi(k\chi^2+(\bar{\bar{b}}-\Omega)\chi))}\mathrm{d}\chi \right|,
\end{aligned}
\tag{79}
$$

where

$$
\bar{\bar{b}} \triangleq b + 2k(t-1)M, \quad U(\chi) = \begin{cases} 1, & 1 \leq \chi \leq M, \\ 0, & \text{others.} \end{cases}
\tag{80}
$$

According to the PSP [30]–[32], (79) can be approximated as

$$
|G(\boldsymbol{G}_t\boldsymbol{h},\Omega)| \approx \left| \sqrt{\frac{-2\pi}{\Phi''(\chi_0,\Omega)}}U(\chi_0) \right| = \begin{cases} \sqrt{\frac{1}{-k}}, & \bar{\bar{b}}+2kM \leq \Omega \leq \bar{\bar{b}}+2k, \\ 0, & \text{others,} \end{cases}
\tag{81}
$$

where $\Phi(\chi,\Omega) \triangleq \pi(k\chi^2 + (\bar{\bar{b}} - \Omega)\chi))$ and $\chi_0 = \frac{\Omega-\bar{\bar{b}}}{2k}$. The relation in (81) 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 calculated as

$$
\Gamma = 1 - \frac{\max_\Omega |G(\boldsymbol{G}_t\boldsymbol{h},\Omega)|}{M} \overset{\text{(a)}}{\leq} 1 - \sqrt{\frac{4r}{\lambda M^2}} \overset{\text{(b)}}{\leq} 1 - \sqrt{\frac{2\sqrt{D^3/\lambda}}{\lambda M^2}} = 1 - \frac{N_{\mathrm{RF}}}{\sqrt[4]{2N}},
\tag{82}
$$

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

$$\Gamma_{\max} = \max\{1 - N_{\mathrm{RF}}/\sqrt[4]{2N}, 0\}, \tag{83}$$

which completes the proof.

APPENDIX B

First of all, we calculate the beam center of $\boldsymbol{G}_t\boldsymbol{\alpha}(N, \Omega, r)$. Based on (81), we can approximate the beam gain of $\boldsymbol{G}_t\boldsymbol{\alpha}(N, \Omega, r)$ as

$$|G(\boldsymbol{G}_t\boldsymbol{\alpha}(N, \Omega, r), \Omega)| = \begin{cases} \sqrt{\frac{1}{-k}}, & \bar{\bar{b}}+2kM \leq \Omega \leq \bar{\bar{b}}+2k, \\ 0, & \text{others}, \end{cases} \tag{84}$$

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

$$B_t = \bar{\bar{b}} + k(M + 1) = \Omega + \frac{\lambda(1 - \Omega^2)(N - (2t - 1)M)}{4r}. \tag{85}$$

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

$$\begin{aligned} \Psi_t &= \frac{r\Omega - \Delta_t\lambda}{\sqrt{r^2 + \Delta_t^2\lambda^2 - 2r\Omega\Delta_t\lambda}} \\ &= \frac{r\Omega - \Omega^2\Delta_t\lambda}{\sqrt{r^2 + \Delta_t^2\lambda^2 - 2r\Omega\Delta_t\lambda}} + \frac{(\Omega^2 - 1)\Delta_t\lambda}{\sqrt{r^2 + \Delta_t^2\lambda^2 - 2r\Omega\Delta_t\lambda}} \\ &\overset{(a)}{\approx} \frac{r\Omega - \Omega^2\Delta_t\lambda}{r - \Omega\Delta_t\lambda} + \frac{(\Omega^2 - 1)\Delta_t\lambda}{r} \\ &= \Omega + \frac{\lambda(1 - \Omega^2)(N - (2t - 1)M)}{4r}, \end{aligned} \tag{86}$$

where the former of (a) holds because we approximate $\sqrt{r^2 + \Delta_t^2\lambda^2 - 2r\Omega\Delta_t\lambda}$ with its one-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 (86) 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 (85) and (86), we have

$$B_t \approx \Psi_t, \tag{87}$$

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] E. Bjornson, L. Van der Perre, S. Buzzi, and E. G. Larsson, "Massive MIMO in sub-6 GHz and mmWave: Physical, practical, and use-case differences," *IEEE Wireless Commun.*, vol. 26, no. 2, pp. 100–108, Apr. 2019.

[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] Z. Qiu, S. Zhou, M. Zhao, and W. Zhou, "Low-complexity precoding by exploiting spatial sparsity in massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 21, no. 7, pp. 4740–4753, Dec. 2022.

[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., early access*, pp. 1–71, 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. 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.

[8] 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.

[9] M. Cui, Z. Wu, Y. Lu, X. Wei, and L. Dai, "Near-field MIMO communications for 6G: Fundamentals, challenges, potentials, and future directions," *IEEE Commun. Mag.*, vol. 61, no. 1, pp. 40–46, Feb. 2023.

[10] C. Feng, W. Shen, J. An, and L. Hanzo, "Weighted sum rate maximization of the mmwave cell-free MIMO downlink relying on hybrid precoding," *IEEE Trans. Wireless Commun.*, vol. 21, no. 4, pp. 2547–2560, Apr. 2022.

[11] X. Sun, C. Qi, and G. Y. Li, "Beam training and allocation for multiuser millimeter wave massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 18, no. 2, pp. 1041–1053, Feb. 2019.

[12] S. Srivastava, R. K. Singh, A. K. Jagannatham, A. Chockalingam, and L. Hanzo, "OTFS transceiver design and sparse doubly-selective CSI estimation in analog and hybrid beamforming aided mmWave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 21, no. 12, pp. 10 902–10 917, July 2022.

[13] 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.

[14] 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, Sep. 2017.

[15] 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.

[16] Z. Zhang, X. Wu, and D. Liu, "Joint precoding and combining design for hybrid beamforming systems with subconnected structure," *IEEE Syst. J.*, vol. 14, no. 1, pp. 184–195, Mar. 2020.

[17] 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.

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

[19] 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.

[20] 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.

[21] 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.

[22] X. Shi, J. Wang, Z. Sun, and J. Song, "Hierarchical codebook-based beam training for extremely large-scale massive MIMO," *arXiv preprint arXiv:2210.03345*, 2022.

[23] 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.

[24] J. A. Zhang, K. Wu, X. Huang, and Y. J. Guo, "Beam alignment for analog arrays based on gaussian approximation," *IEEE Trans. Veh. Technol., early access*, pp. 1–6, 2023.

[25] 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.

[26] 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.

[27] 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.

[28] 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.

[29] 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.

[30] C. M. Bender and S. A. Orszag, *Advanced Mathematical Methods for Scientists and Engineers*. New York, NY, USA: McGraw-Hill, 1978.

[31] F. W. J. Olver, *Asymptotics and Special Functions*. Boca Raton, FL, USA: CRC Press, 1997.

[32] N. M. Temme, *Asymptotic Methods for Integrals*. Singapore: World Scientific, vol. 11, 2014.

[33] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Cliffs, NJ, USA: Prentice-Hall, 1998.

[34] M. A. Richards, *Fundamentals of Radar Signal Processing*, 2nd ed. New York, NY, USA: McGraw-Hill, 2005.

[35] 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.

[36] 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.

[37] H. Wang, R. Lu, Z. Peng, and M. Li, "Timestamp-free clock parameters tracking using extended Kalman filtering in wireless sensor networks," *IEEE Trans. Commun.*, vol. 69, no. 10, pp. 6926–6938, Oct. 2021.

[38] 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.