Two-Stage Hybrid-Field Beam Training for Ultra-Massive MIMO Systems

Kangjian Chen*, Chenhao Qi* and Cheng-Xiang Wang*†‡
*School of Information Science and Engineering, Southeast University, Nanjing, China
†National Mobile Communications Research Laboratory, Southeast University
‡Purple Mountain Laboratories, Nanjing, China
Email: {kjchen, qch, chxwang}@seu.edu.cn

2022 IEEE/CIC International Conference on Communications in China (ICCC) | 978-1-6654-8480-0/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICCC55456.2022.9880793

***Abstract*—In this paper, by considering both the near field and far field, the beam training is investigated for an ultra-massive multiple-input-multiple-output system with a partially-connected hybrid combing structure. As the Rayleigh distance decreases quadratically with the reduction of the antenna number, the near-field effect of subarrays is much weaker than that of the whole array. Motivated by this, far-field channel steering vectors of a subarray are used to approximate its near-field channel steering vectors. Then a two-stage hybrid-field beam training scheme that works for both the near field and far field is proposed. In the first stage, each subarray independently uses multiple far-field channel steering vectors for analog combining. In the second stage, for each codeword in a predefined hybrid-field codebook, a dedicated digital combiner is designed to combine the output of the analog combiner from the first stage. Then, from the hybrid-field codebook, the codeword corresponding to the dedicated digital combiner that achieves the largest combining power is selected. Note that the dedicated digital combiners can be obtained offline before the beam training and the combining power can be computed in parallel. Simulation results show that the proposed scheme can approach the performance of the hybrid-field beam sweeping but with considerable reduction in training overhead.**

***Index Terms*—Beam training, hybrid combining, near field, ultra-massive MIMO (UM-MIMO).**

I. Introduction

Massive multiple-input-multiple-output (MIMO) is a key technology for the fifth generation (5G) wireless communications [1], [2]. Equipped with large antenna arrays, the base station (BS) can improve the spectral efficiency dramatically by MIMO beamforming to exploit the spatial degree of freedom (DoF). For the future sixth generation (6G) wireless communications, ultra-massive MIMO (UM-MIMO) with much more antennas than the existing massive MIMO is considered as a key technology to further improve the spectral efficiency [3], [4].

Due to large power consumption, each antenna with a dedicated radio frequency (RF) chain is impractical for UM-MIMO [5]. Consequently, hybrid structure, where a small number of RF chains are connected to a large number of antennas, is developed for UM-MIMO systems [6]. 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 extend to different sizes of antennas in blocks [5].

One important difference between UM-MIMO and the existing massive MIMO is that the channel features of the former differ from that of the latter. For UM-MIMO, according to 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 [6]. On one hand, the Rayleigh distance increases linearly with the increase of the wavelength. On the other hand, when fixing the wavelength, the Rayleigh distance increases quadratically with the increase of the number of antennas. As a result, the far field assumption in existing massive MIMO may not hold for the UM-MIMO systems, especially when the user is close to the BS. In this context, channel state information (CSI) acquisition methods considering both the near-filed effect and far-field effect are needed for UM-MIMO. The main stream CSI acquisition methods can be categorized into channel estimation and beam training [7]. The channel estimation methods usually focus 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 [6], to estimate the near-field channels in UM-MIMO, a polar-domain simultaneous orthogonal matching pursuit (P-SOMP) algorithm is proposed. However, it only considers the random beamforming instead of the directional beamforming, where the random beamforming cannot offer enough beamforming gain and may limit the signal coverage of the BS as well as degrading the channel estimation performance at low SNRs. To the best knowledge of the authors, so far there has been no literature working on efficient beam training for the UM-MIMO system with near-field effect.

In this paper, by taking into account both the near field and far field, we investigate the beam training for a UM-MIMO system with a partially-connected hybrid combing structure. As the Rayleigh distance decreases quadratically with the reduction of the antenna number, the near-field effect

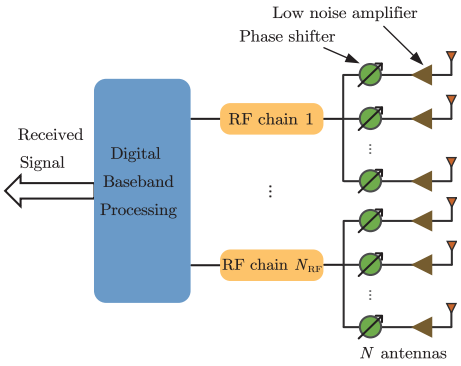

Fig. 1. Illustration of a partially-connected hybrid combing structure.

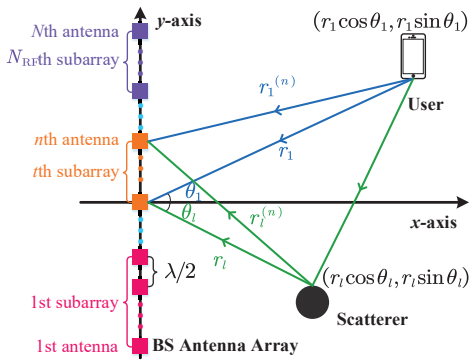

Fig. 2. Illustration of a multipath channel model.

of subarrays is much weaker than that of the whole array. Motivated by this, we use far-field channel steering vectors of a subarray to approximate its near-field channel steering vectors. Then we propose a two-stage hybrid-field beam training scheme that works for both the near field and far field. In the first stage, each subarray independently uses multiple far-field channel steering vectors for analog combining. In the second stage, for each codeword in a predefined hybrid-field codebook, we design a dedicated digital combiner, which is used to combine the output of the analog combiner from the first stage. Then, from the hybrid-field codebook, we select the codeword corresponding to the dedicated digital combiner that achieves the largest combining power. Note that the dedicated digital combiners can be obtained offline before the beam training and the combining power can be computed in parallel.

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 number. 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.

II. System Model

As shown in Fig. 1, we consider uplink beam training for a UM-MIMO system including a user and a BS. At the BS, a large uniform linear array (ULA) of N antennas with half wavelength interval and a partially-connected hybrid combining structure with N_{RF} RF chains are employed, where the hybrid combining includes analog combining and digital combining. In practice, the BS with hybrid structure has much 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 the analog combining and digital combining at the BS side, while a single-antenna user is considered for simplification.

During the uplink beam training, the signal transmitted by the user is denoted as $x_k, k = 1, 2, \ldots, K$, where K 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

$$y_k = \boldsymbol{v}_k \boldsymbol{W}_k \boldsymbol{h} x_k + \boldsymbol{v}_k \boldsymbol{W}_k \boldsymbol{\eta} \tag{1}$$

where $\boldsymbol{W}_k \in \mathbb{C}^{N_{\mathrm{RF}} \times N}$ is the analog combiner, $\boldsymbol{v}_k \in \mathbb{C}^{1 \times N_{\mathrm{RF}}}$ is the digital combiner, and $\boldsymbol{\eta} \in \mathbb{C}^N$ is an additive white Gaussian noise 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 $\boldsymbol{v}_k$, 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 by existing literature according to the distance between the user and the BS [6]. 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 termed as the far-filed channel; otherwise, the wireless channel is termed as the near-filed channel. Since a half-wavelength-interval ULA is adopted in this work, the array aperture of the BS is $D = N\lambda/2$. In the following, we describe both the far-filed channel and the near-field channel based on a multipath channel model, and then propose an uplink beam training method to estimate the multipath channel.

As shown in Fig. 2, a multipath 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 coordinate 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 $\boldsymbol{p}_l$ for $l \geq 1$ and the nth antenna can be expressed as

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

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

$$\boldsymbol{h} = \sqrt{\frac{N}{L}} \sum_{l=1}^{L} g_l e^{-\frac{j2\pi r_l}{\lambda}} \boldsymbol{\alpha}(N, \Omega_l, r_l) \tag{4}$$

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

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

Note that the channel steering vector in (5) can be used to describe both the far-field channel and the near-field channel. If $r_l > Z$, then $r_l^{(n)}$ 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 $\sqrt{1+\epsilon} \approx 1 + \frac{1}{2}\epsilon$. As a result, (5) can be approximated by

$$\begin{aligned}\boldsymbol{\alpha}(N, \Omega_l, r_l) &\approx \frac{1}{\sqrt{N}} \left[e^{j2\pi\Omega_l\delta_1}, e^{j2\pi\Omega_l\delta_2}, \cdots, e^{j2\pi\Omega_l\delta_N}\right]^{\mathrm{T}} \\ &\triangleq \boldsymbol{\beta}(N, \Omega_l).\end{aligned} \tag{6}$$

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 (6) is not accurate enough. Therefore, we use $\boldsymbol{\alpha}(N, \Omega_l, r_l)$ to represent the near-field channel steering vector.

To estimate the multipath channel, codebook-based beam training is widely adopted [8]. Since the user is either in the far field or near field, a hybrid-field codebook considering both the near-field effect and the far-field effect will be developed in the following. First, we establish a far-field codebook $\boldsymbol{C}_{\mathrm{f}} \in \mathbb{C}^{N\times N}$, based on the far-field channel steering vectors [2], where the nth column of $\boldsymbol{C}_{\mathrm{f}}$ is denoted as

$$[\boldsymbol{C}_{\mathrm{f}}]_{:,n} \triangleq \boldsymbol{\beta}\left(N, \frac{2n-1-N}{N}\right) \tag{7}$$

for $n = 1, 2, \ldots, N$. Then we establish a near-field codebook $\boldsymbol{C}_{\mathrm{n}} \in \mathbb{C}^{N\times(NS)}$. Since the near-field channel is relevant to both the distance and the angle, we quantize the distance and and the angle by S samples and N samples, respectively. The nth angle sample is $\Theta_n = (2n-1-N)/N$, $n = 1, 2, \cdots, N$. The sth distance sample is

$$d_{n,s} = \frac{N^2\lambda(1-\Theta_n^2)}{8s} \tag{8}$$

for $s = 1, 2, \cdots, S$ [6]. Then $\boldsymbol{C}_{\mathrm{n}}$ can be designed as

$$\boldsymbol{C}_{\mathrm{n}} \triangleq \{\boldsymbol{C}_1, \boldsymbol{C}_2, \cdots, \boldsymbol{C}_N\} \tag{9}$$

where the sth column of $[\boldsymbol{C}_n]$ is $[\boldsymbol{C}_n]_{:,s} = \boldsymbol{\alpha}(N, \Theta_n, d_{n,s})$. Then the hybrid-field codebook is defined as

$$\boldsymbol{C}_{\mathrm{h}} \triangleq \{\boldsymbol{C}_{\mathrm{n}}, \boldsymbol{C}_{\mathrm{f}}\} \in \mathbb{C}^{N\times(NS+N)}. \tag{10}$$

Based on (1), we have

$$y_k = [\boldsymbol{C}_{\mathrm{h}}]_{:,k}^{\mathrm{H}}\boldsymbol{h}x_k + [\boldsymbol{C}_{\mathrm{h}}]_{:,k}^{\mathrm{H}}\boldsymbol{\eta} \tag{11}$$

for $k = 1, 2, \cdots, NS+N$. The beam training aims at finding the codeword in $\boldsymbol{C}_{\mathrm{h}}$ best fit for the multipath channel, which can be expressed as

$$\widetilde{k} = \arg \max_{k=1,2,\cdots,NS+N} \left|[\boldsymbol{C}_{\mathrm{h}}]_{:,k}^{\mathrm{H}}\boldsymbol{h}\right|. \tag{12}$$

To solve (12), we need to test all the codewords in $\boldsymbol{C}_{\mathrm{h}}$ one by one. We term this method as the hybrid-field beam sweeping. To perform the hybrid-field beam sweeping, $(NS+N)$ times of beam training are needed. Comparing (6) with (5), we observe that larger overhead is needed by the near-field beam training than the far-field beam training, because the former needs S times beam training for different distance 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 in the following section.

III. Two-Stage Beam Training

In this section, we will use far-field channel steering vectors of a subarray to approximate its near-field channel steering vectors. Then we propose a two-stage hybrid-field beam training scheme that works for both the near field and far field.

We define an analog combiner as

$$\boldsymbol{W} \triangleq \mathrm{blkdiag}\{\boldsymbol{w}_1, \boldsymbol{w}_2, \cdots, \boldsymbol{w}_{N_{\mathrm{RF}}}\} \tag{13}$$

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

$$\max_{\boldsymbol{W},\boldsymbol{v}} \quad \left|\boldsymbol{v}\boldsymbol{W}\boldsymbol{h}\right| \tag{14a}$$

$$\text{s.t.} \quad \|\boldsymbol{v}\boldsymbol{W}\|_2 = 1, \left|[\boldsymbol{w}_t]_m\right| = 1, \quad m = 1, 2, \ldots, M,\ t = 1, 2, \ldots, N_{\mathrm{RF}}. \tag{14b}$$

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 (14) can be rewritten as

$$\max_{\boldsymbol{W},\boldsymbol{v}} \left|\boldsymbol{v}\boldsymbol{W}\boldsymbol{\alpha}(N, \Omega, r)\right| \tag{15a}$$

$$\text{s.t.} \quad \|\boldsymbol{v}\boldsymbol{W}\|_2 = 1, \left|[\boldsymbol{w}_t]_m\right| = 1, \quad m = 1, 2, \ldots, M,\ t = 1, 2, \ldots, N_{\mathrm{RF}} \tag{15b}$$

where we omit the subscript of Ω and r in (15a) to simplify the notation. According to the Cauchy-Schwartz inequality, we have

$$\left|\boldsymbol{v}\boldsymbol{W}\boldsymbol{\alpha}(N, \Omega, r)\right| \leq \|\boldsymbol{v}\|_2\ \|\boldsymbol{W}\boldsymbol{\alpha}(N, \Omega, r)\|_2. \tag{16}$$

The equality of (16) 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 to each other, we can achieve the equality of (16). Consequently, we may first design $\boldsymbol{W}$ and then design $\boldsymbol{v}$.

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

$$\max_{\boldsymbol{W}} \ \big\|\boldsymbol{W}\boldsymbol{\alpha}(N,\Omega,r)\big\|_2 \tag{17a}$$

$$\text{s.t.} \ \ \big|[\boldsymbol{w}_t]_m\big| = 1, \tag{17b}$$

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

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 of (17) is divided into N_{RF} independent subproblems with the tth subproblem expressed as

$$\max_{\boldsymbol{w}_t} \ \big|\boldsymbol{w}_t\boldsymbol{G}_t\boldsymbol{\alpha}(N,\Omega,r)\big| \tag{18a}$$

$$\text{s.t.} \ \ \big|[\boldsymbol{w}_t]_m\big| = 1, m = 1,2,\ldots,M \tag{18b}$$

where $\boldsymbol{G}_t \triangleq [\boldsymbol{0}_{M\times(t-1)M}, \boldsymbol{I}_M, \boldsymbol{0}_{M\times(N_{\mathrm{RF}}-t)M}]$. The optimal solution of (18) is

$$\overline{\boldsymbol{w}}_t = \sqrt{N}\big(\boldsymbol{G}_t\boldsymbol{\alpha}(N,\Omega,r)\big)^{\mathrm{H}}, \ t = 1,2,\ldots,N_{\mathrm{RF}}. \tag{19}$$

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 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 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. As a results, for each subarray, we use far-field channel steering vectors expressed as

$$\widehat{\boldsymbol{w}}_t = \sqrt{M}\boldsymbol{\beta}(M,\hat{\Psi}_t)^{\mathrm{H}}, \ t = 1,2,\ldots,N_{\mathrm{RF}} \tag{20}$$

to approximate the near-field channel steering vectors in (19), so that we can use (20) for both the far-field and near-field channels. Note that the aforementioned approximation is similar to that in (6). In (20), $\hat{\Psi}_t$ represents the sine result of the angle that points from the center of the tth subarray to the user. 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 $\hat{\Psi}_t$ can be computed as

$$\hat{\Psi}_t = \begin{cases} \frac{r\Omega - \Delta_t\lambda}{\sqrt{r^2+\Delta_t^2\lambda^2-2r\Omega\Delta_t\lambda}}, & r \le Z, \\ \Omega \qquad\qquad\quad , & r > Z. \end{cases} \tag{21}$$

Note that $\hat{\Psi}_t$ is different for different subarrays in the near field, where such difference is more apparent if the user is closer to the BS. For example, if $N = 256$, $\lambda = 0.003$ m, $N_{\mathrm{RF}} = 4$, $r = 5$ m and $\Omega = 0$, we can obtain $\hat{\Psi}_1 \approx 0.0288$, $\hat{\Psi}_2 \approx 0.0096$, $\hat{\Psi}_3 \approx -0.0096$ and $\hat{\Psi}_4 \approx -0.0288$ via (21). It is seen that $\hat{\Psi}_4 - \hat{\Psi}_1 \approx 0.0576$, but the beam coverage of a subarray channel steering vector is only $2/64 \approx 0.0312$. In this context, if the same sine angle is assigned to different subarrays for analog combining in the near field, there will be severe loss of beamforming gain. Therefore, we need to compute $\hat{\Psi}_t$ for different t.

By substituting (20) into (13), we write the designed analog combiner for (15) as

$$\widehat{\boldsymbol{W}} = \mathrm{blkdiag}\{\widehat{\boldsymbol{w}}_1, \widehat{\boldsymbol{w}}_2, \cdots, \widehat{\boldsymbol{w}}_{N_{\mathrm{RF}}}\}. \tag{22}$$

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 (15) can be expressed as

$$\max_{\boldsymbol{v}} \ \boldsymbol{v}\widehat{\boldsymbol{W}}\boldsymbol{\alpha}(N,\Omega,r) \tag{23a}$$

$$\text{s.t.} \ \ \|\boldsymbol{v}\|_2 = 1/\sqrt{N}. \tag{23b}$$

Note that (15a) can be rewritten as (23a) 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 (23) is

$$\widehat{\boldsymbol{v}} = \frac{\Big(\widehat{\boldsymbol{W}}\boldsymbol{\alpha}(N,\Omega,r)\Big)^{\mathrm{H}}}{\sqrt{N}\big\|\widehat{\boldsymbol{W}}\boldsymbol{\alpha}(N,\Omega,r)\big\|_2}. \tag{24}$$

For each subarray, the commonly-used DFT codebook for beam training is

$$\boldsymbol{\Phi} = \{\sqrt{M}\boldsymbol{\beta}(M,\Phi_1), \sqrt{M}\boldsymbol{\beta}(M,\Phi_2), \cdots, \sqrt{M}\boldsymbol{\beta}(M,\Phi_M)\}. \tag{25}$$

where

$$\Phi_m = (2m-1-M)/M, \ m = 1,2,\ldots,M. \tag{26}$$

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 fit for the channel is

$$\widetilde{m}_t = \arg\min_{m=1,2,\cdots,M} |\Phi_m - \hat{\Psi}_t|. \tag{27}$$

Therefore, the designed analog combiner and digital combiner according to (22) and (24), respectively, are

$$\widetilde{\boldsymbol{W}} = \mathrm{blkdiag}\{\widetilde{\boldsymbol{w}}_1, \widetilde{\boldsymbol{w}}_2, \cdots, \widetilde{\boldsymbol{w}}_{N_{\mathrm{RF}}}\}. \tag{28}$$

$$\widetilde{\boldsymbol{v}} = \frac{\Big(\widetilde{\boldsymbol{W}}\boldsymbol{\alpha}(N,\Omega,r)\Big)^{\mathrm{H}}}{\sqrt{N}\big\|\widetilde{\boldsymbol{W}}\boldsymbol{\alpha}(N,\Omega,r)\big\|_2}. \tag{29}$$

where

$$\widetilde{\boldsymbol{w}}_t = [\boldsymbol{\Phi}]^{\mathrm{H}}_{:,\widetilde{m}_t}, \ t = 1,2,,\ldots,N_{\mathrm{RF}}. \tag{30}$$

based on (20).

For the partially-connected structure, each subarray is solely connected to a 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 the user is near-filed 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 kth combiner according to (25) is

$$\overline{\boldsymbol{W}}_k = \mathrm{blkdiag}\{\sqrt{M}\boldsymbol{\beta}(M,\Phi_k)^{\mathrm{H}},\cdots,\sqrt{M}\boldsymbol{\beta}(M,\Phi_k)^{\mathrm{H}}\}. \tag{31}$$

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

$$\boldsymbol{z}_k = \overline{\boldsymbol{W}}_k\boldsymbol{h}x_k + \overline{\boldsymbol{W}}_k\boldsymbol{\eta},\ k=1,2,\ldots,M. \tag{32}$$

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 $p=1,2,\cdots,NS+N$, 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 (10), we have

$$[\boldsymbol{C}_{\mathrm{h}}]_{:,p} = \begin{cases} \boldsymbol{\alpha}\left(N,\Theta_{\bar{n}},d_{\bar{n},\bar{s}}\right), & p\le NS, \\ \boldsymbol{\beta}\left(N,\frac{2(p-NS)-1-N}{N}\right), & p>NS \end{cases} \tag{33}$$

where

$$\bar{n}=\lfloor p/S\rfloor,\ \bar{s}=p-(\bar{n}-1)S. \tag{34}$$

Suppose $[\boldsymbol{C}_{\mathrm{h}}]_{:,p}$ is the codeword best fit for the channel. We will design analog combiner $\boldsymbol{F}_p$ and digital combiner $\widetilde{\boldsymbol{v}}_p$ so that $\widetilde{\boldsymbol{v}}_p\boldsymbol{F}_p$ achieves the similar beam gain as $[\boldsymbol{C}_{\mathrm{h}}]_{:,p}$. Replacing $\boldsymbol{\alpha}(N,\Omega,r)$ in (15) by $[\boldsymbol{C}_{\mathrm{h}}]_{:,p}$, we can obtain $\widetilde{\boldsymbol{w}}_t^{(p)}$ via (30). Similar to (28) and (29), we design the analog combiner and digital combiner respectively as

$$\boldsymbol{F}_p = \mathrm{blkdiag}\{\widetilde{\boldsymbol{w}}_1^{(p)},\widetilde{\boldsymbol{w}}_2^{(p)},\cdots,\widetilde{\boldsymbol{w}}_{N_{\mathrm{RF}}}^{(p)}\}, \tag{35}$$

$$\widetilde{\boldsymbol{v}}_p = \frac{\left(\boldsymbol{F}_p[\boldsymbol{C}_{\mathrm{h}}]_{:,p}\right)^{\mathrm{H}}}{\sqrt{N}\left\|\boldsymbol{F}_p[\boldsymbol{C}_{\mathrm{h}}]_{:,p}\right\|_2}. \tag{36}$$

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 define

$$\widetilde{\boldsymbol{z}}_p = \boldsymbol{F}_p\boldsymbol{h}x_p + \boldsymbol{F}_p\boldsymbol{\eta}, \tag{37}$$

$$\widetilde{y}_p = \widetilde{\boldsymbol{v}}_p\widetilde{\boldsymbol{z}}_p,\ p=1,2,\cdots,NS+N. \tag{38}$$

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 (32), because both $\boldsymbol{F}_p$ and $\overline{\boldsymbol{W}}_k$ are composed of channel steering vectors from the same set $\boldsymbol{\Phi}$ in (25). We can obtain $\widetilde{\boldsymbol{z}}_p$ by setting

$$[\widetilde{\boldsymbol{z}}_p]_t = [\boldsymbol{z}_{\widetilde{m}_t}]_t \tag{39}$$

where $\widetilde{m}_t$ can be obtained from (27) during the design of $\boldsymbol{F}_p$ in (35). Note that (38) can be computed in parallel to speed up the beam training. From $\{\widetilde{y}_1,\widetilde{y}_2,\ldots,\widetilde{y}_{NS+N}\}$, we select one with the largest power, which can be expressed as

$$\widetilde{p} = \arg\max_{p=1,2,\cdots,NS+N}\left|\widetilde{y}_p\right|^2. \tag{40}$$

Finally, 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 $|\widetilde{y}_{\widetilde{p}}|^2$. The detailed steps of the proposed two-stage hybrid-field beam training scheme are summarized in **Algorithm 1**.

Algorithm 1 Two-Stage Hybird-Field Beam Training

1: **Input:** N, N_{RF}, M, S, λ.
2: **First Stage:**
3: Obtain $\boldsymbol{z}_k,\ k=1,2,\ldots,M$ via (32).
4: **Second Stage:**
5: **for** $p=1,2,\cdots NS+N$ **do**
6: Obtain $[\boldsymbol{C}_{\mathrm{h}}]_{:,p}$ via (33).
7: Obtain $\boldsymbol{F}_p$ and $\widetilde{\boldsymbol{v}}_p$ via (35) and (36), respectively.
8: Obtain $\widetilde{\boldsymbol{z}}_p$ via (39).
9: Obtain $\widetilde{y}_p$ via (38).
10: **end for**
11: Obtain $\widetilde{p}$ via (40).
12: **Output:** $[\boldsymbol{C}_{\mathrm{h}}]_{:,\widetilde{p}}$.

Compared to the hybrid-field beam sweeping based on $\boldsymbol{C}_{\mathrm{h}}$ in (10) that needs $NS+N$ times of beam training, the training overhead of the two-stage hybrid-field beam training scheme is substantially reduced to M. The computational complexity of the proposed scheme mainly comes from step 10 in **Algorithm 1** and is totally $\mathcal{O}(N_{\mathrm{RF}}N(S+1))$.

IV. Simulation Results

Now we evaluate the performance of the proposed two-stage hybrid-field beam training scheme. We consider a UM-MIMO system equipped with $N=256$ antennas. The antenna array is composed of $N_{\mathrm{RF}}=4$ subarrays with each subarray having $M=64$ 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 $g_2\sim\mathcal{CN}(0,0.1)$ and $g_3\sim\mathcal{CN}(0,0.1)$. The channel angle Ω_l of the lth path obeys the uniform distribution between $[-\sqrt{3}/2,\sqrt{3}/2]$. We set $S=6$ for the hybrid-field codebook $\boldsymbol{C}_{\mathrm{h}}$. During the uplink beam training, the length of the signal transmitted by the user is $K=64$, which is also set to be the pilot length of P-SOMP for fair comparisons.

In Fig. 3, we compare the proposed two-stage hybrid-field beam training scheme with P-SOMP [6], the hybrid-field beam sweeping and the far-filed beam sweeping [9] in terms of spectral efficiency. The distances between the BS and the user or scatterers obey the uniform distribution between $[5,10]$ m. From Fig. 3, the hybrid-field beam sweeping can achieve better performance than the other three schemes,

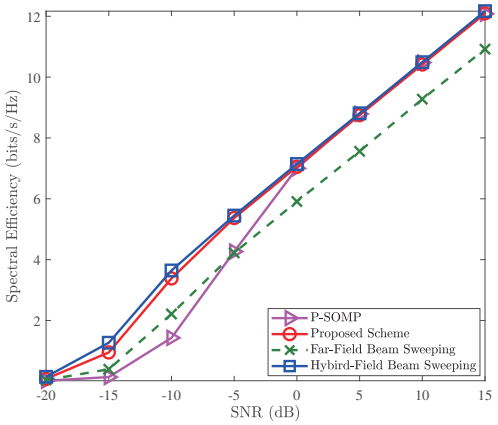

Fig. 3. Comparisons of the spectral efficiency for different methods.

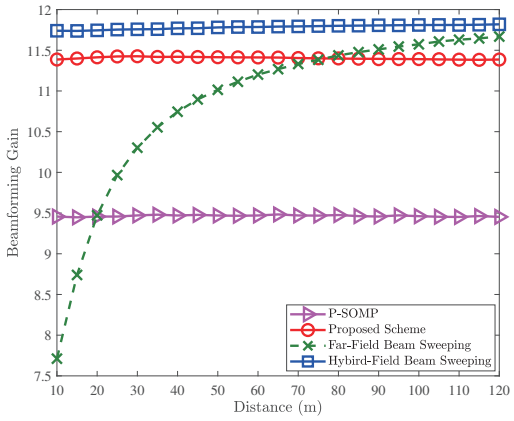

Fig. 4. Comparisons of the beamforming gains for different methods.

which lies in the fact that hybrid-field beam sweeping exhaustively tests all the codewords in $\boldsymbol{C}_{\mathrm{h}}$ and needs much more times of beam training than the other three schemes. The performance of P-SOMP is worse than that of the other three schemes at low SNRs, such as -5 dB, because the random beamformings of P-SOMP cannot achieve enough beamforming gain and significantly degrade the performance. The performance of the far-field beam sweeping is worse than that of the other three schemes at high SNRs because beamforming gain of the far-field channel steering vector will decrease in the near field. Most importantly, the performance of the proposed scheme can approach the performance of the hybrid-field beam sweeping at various SNR conditions.

In Fig. 4, we compare the proposed two-stage hybrid-field beam training scheme with P-SOMP, the hybrid-field beam sweeping and the far-filed beam sweeping in terms of beamforming gain. The distances between the BS and the user or scatterers obey the uniform distribution between $[5, r]$ m, where r ranges from 10 to 120. The SNR is fixed to be -5 dB. From Fig. 4, the hybrid-field beam sweeping achieves the highest beam gain at different distances. 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 three schemes are robust to the distance because they consider both the near-filed channel and the far-field channel. In particular, the performance of the proposed scheme can approach that of the hybrid-field beam sweeping with only slight loss of beamforming gain at different distances.

We also compare the training overhead of different schemes. The training overhead of the hybrid-field beam sweeping, the far-field beam sweeping, P-SOMP and the proposed scheme are $N(S+1)$, N, K and M, respectively. Under the simulation setting, these four schemes require 1792, 256, 64 and 64 time slots, respectively, where the proposed scheme can approach the performance of the hybrid-field beam sweeping with 96.43% reduction in training overhead.

V. Conclusion

In this paper, we have proposed a two-stage hybrid-field beam training scheme for UM-MIMO systems with partially-connected hybrid combining structure. Future work will be continued with the focus on efficient beam training for UM-MIMO.

Acknowledgment

This work is supported in part by National Natural Science Foundation of China (NSFC) under Grant 62071116 and 61960206006, and by National Key Research and Development Program of China under Grant 2021YFB2900404 and 2020YFB1807205.

References

[1] 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.

[2] 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.

[3] X. You, C. 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.

[4] 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.

[5] 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.

[6] 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.

[7] 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.

[8] K. Chen, C. Qi, and G. Y. Li, "Two-step codeword design for millimeter wave massive MIMO systems with quantized phase shifters," *IEEE Trans. Signal Process.*, vol. 68, no. 1, pp. 170–180, Jan. 2020.

[9] 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.