Hierarchical Beam Training for Extremely Large-Scale MIMO: From Far-Field to Near-Field

Yu Lu, *Student Member, IEEE*, Zijian Zhang, *Student Member, IEEE*, and Linglong Dai, *Fellow, IEEE*

***Abstract*—Extremely large-scale MIMO (XL-MIMO) is a promising technique for future 6G communications. The sharp increase in the number of antennas results in a transition of electromagnetic propagation from the far-field to the near-field. Due to the near-field effect, the exhaustive near-field beam training at all angles and distances requires very high overhead. The improved fast near-field beam training scheme based on time-delay structure can reduce the overhead, but it suffers from very high hardware costs and energy consumption caused by time-delay circuits. In this paper, we propose a near-field two dimension (2D) hierarchical beam training scheme to reduce the overhead without the need for extra hardware circuits. Specifically, we first formulate the multi-resolution near-field codewords design problem covering different angle and distance coverages. Next, inspired by phase retrieval problems in digital holography imaging technology, we propose a Gerchberg-Saxton (GS)-based algorithm to acquire the theoretical codeword by considering the fully digital architecture. Based on the theoretical codeword, an alternating optimization algorithm is proposed to acquire the practical codeword considering the hybrid digital-analog architecture. Finally, with the designed multi-resolution codebooks, we propose a near-field 2D hierarchical beam training scheme to significantly reduce the training overhead, which is verified by extensive simulation results.**

***Index Terms*—Extremely large-scale MIMO (XL-MIMO), extremely large-scale antenna array (ELAA), beam training, codebook design.**

I. Introduction

With the development of new applications such as digital twins, 6G is predicted to achieve a 10-fold increase in spectrum efficiency than 5G [1], [2]. The extremely large-scale MIMO (XL-MIMO) is a technology with bright prospects for 6G to achieve ultra-high spectrum efficiency [3], [4]. In XL-MIMO systems, the base station (BS) usually deploys an extremely large-scale antenna array (ELAA), which consists of hundreds or even thousands of antennas. ELAA in the XL-MIMO system is expected to drastically improve spatial resolution to realize a high spatial multiplexing gain in 6G. In order to obtain spatial multiplexing gain, XL-MIMO should generate a directional beam with high array gain by beamforming.

To support beamforming, beam training should be conducted to search the optimal beamforming vector, i.e., codeword, in the predefined codebook. As the number of BS antennas in XL-MIMO systems is much larger than that of 5G systems, the high-dimensional XL-MIMO beam training overhead will be overwhelming.

All authors are with the Department of Electronic Engineering, Tsinghua University as well as Beijing National Research Center for Information Science and Technology (BNRist), Beijing 100084, China (e-mails: y-lu19@mails.tsinghua.edu.cn, zhangzj20@tsinghua.edu.cn, daill@tsinghua.edu.cn).

This work was supported in part by the National Key Research and Development Program of China (Grant No. 2020YFB1807201) and in part by the National Natural Science Foundation of China (Grant No. 62031019).

A. Prior Works

There are two typical categories of beam training methods for MIMO, which are far-field beam training and near-field beam training respectively. For the first category, since the antenna number at BS is usually not very large in 3G-5G systems, the MIMO channel is modeled considering the planar wave assumption in the far-field region, where the array response vector of the far-field channel is only associated with the angle. In this case, the orthogonal Discrete Fourier Transform (DFT) codebook can be utilized in beam training to capture the physical angle information in the angle-domain of the channel paths [5], [6]. However, since the size of the DFT codebook is proportional to the number of antennas at BS, the DFT codebook suffers from very high training overhead when it comes to XL-MIMO systems. Thus, to reduce the beam training overhead, some far-field hierarchical beam training schemes were proposed [7], [8]. The fundamental concept of hierarchical beam training involves systematically searching from the lowest-resolution codebook to the highest-resolution codebook, moving through layers one at a time. This sequential approach results in a gradual reduction of the required angle range that needs to be scanned at each layer. The challenge in hierarchical beam training lies in how to design codebooks with different resolutions [9]. The design criterion for each codeword in the codebook is to maximize the gain at the target angle while minimizing the gain at other angles. This is particularly critical for low-resolution codebooks, where the corresponding beams are broad, potentially causing energy leakage at non-target angles and thereby preventing accurate angle information acquisition during beam training. With the help of multi-resolution codebooks, the overhead of the hierarchical beam training becomes proportional to the logarithm of the antenna number at BS [8].

As the antenna number dramatically increases in 6G XL-MIMO systems, the near-field range expands by orders of magnitude, which can extend to several hundred meters [10]. Thus, the XL-MIMO channel should be modeled with the spherical wave assumption in the near-field region instead of the far-field region. In this case, the existing far-field beam training schemes may not be valid for the near-field XL-MIMO channel. To cope with this problem, near-field beam training should be utilized to match the near-field XL-MIMO channel feature. For the second category, i.e., near-field beam training, the array response vector of the near-field channel is not only associated with the angle but also to distance. Thus, to capture both the angle and distance information of the channel paths accurately, it is advisable to employ a polar-domain codebook [11] rather than a DFT codebook. In particular, the DFT codebook divides the physical angle

space into multiple grids, and from each of these grids, a far-field array response vector is selected as a codeword within the DFT codebook. This allows for the utilization of the angle information associated with a far-field path. Inspired by the DFT codebook in angle-domain, the polar-domain codebook [11] divides the entire "angle-distance" space into multiple grids. From each grid, we sample a near-field array response vector as a codeword to construct a new codebook. The reason why this new codebook is named as polar-domain codebook is that "angle-distance" space can also be regarded as the polar coordinate space. Accordingly, the size of the polar-domain codebook is the product of the antenna number at BS and the number of sampled distance grids. Since only one angle and one distance can be measured in each time slot, the exhaustive search method for near-field beam training has a very high overhead [12]. To reduce the overhead, the authors of [13] introduce a two-phase fast near-field beam training method compared to exhaustive search methods. The first phase employs far-field codebooks and angle-domain beam sweeping to determine candidate angles, while the second phase uses a polar-domain codebook to find the best effective distance for the user. Moreover, the deep learning based near-field beam training scheme is proposed in [14], where the received signals from far-field wide beams are used to estimate the optimal near-field beam by neural networks to reduce the beam training overhead. Nevertheless, the overhead of the above schemes is proportional linearly to the number of antennas and is still large for the XL-MIMO system. To address this problem, we have proposed a fast time-delay based near-field beam training for XL-MIMO with low overhead [15], where each antenna is connected to one time-delay circuit implemented by digital delay circuits [15] or analog baseband delay lines [16]. Each time-delay circuit is able to adjust the frequency-dependent phase shifts on wideband signals by manipulating the delay in each antenna branch. In this case, near-field beams can be flexibly controlled by extra time-delay hardware circuits and then focus on different angles and distances at different frequencies. Thus, the near-field beams with different frequencies can radiate towards different directions to realize the fast beam training by searching multiple directions in one time slot. However, the time-delay based beamforming structure will lead to not only high hardware costs but also very high energy consumption [15], [16], especially for XL-MIMO systems with a large number of antennas.

B. Contributions

Thus, to design a general and low-overhead beam training scheme, we propose a near-field two dimension (2D) hierarchical beam training scheme by designing the multi-resolution codebooks referring to the hierarchical beam training in the far-field scenario. Our contributions are summarized as follows.

1) We first formulate the problem of near-field codeword design. Specifically, compared with the far-field case, the ideal beam pattern of near-field codeword should not only cover a certain angle range but also a certain distance range. By considering ideal fully digital architecture, we provide the design problem of the near-field theoretical codeword. Then, based on the theoretical codeword, we formulate the problem of a practical codeword with assumptions of the hybrid digital-analog structure and quantized phase shifts in practice.
2) In order to design the near-field theoretical codeword, inspired by the Gerchberg–Saxton (GS) algorithm in phase retrieval problems for digital holography imaging, we propose a GS-based theoretical codeword design algorithm for a fully digital architecture. Different from the original GS algorithm, we adapt the transformation methods from Fourier transform to polar-domain transform to match the near-field assumption. Moreover, the power constraint instead of amplitude measurements is considered in each iteration to control the power of the codeword.
3) Since fully digital architecture with high energy assumption is not available in a practical XL-MIMO system, we then design the practical codeword considering the hybrid digital-analog architecture. Based on the theoretical codeword, an alternating optimization algorithm is proposed to acquire the practical codeword, where the digital beamforming vector and the analog beamforming matrix are optimized iteratively. Specifically, in each iteration, the digital beamforming vector is obtained by a closed-form solution. Meanwhile, phases of the entries in the analog beamforming matrix are solved individually by a high-efficient iterative search method.
4) Next, we generate multi-resolution codebooks based on the practical codewords obtained by the alternating optimization algorithm. With the aid of multi-resolution codebooks with different angle coverages and distance coverages, we propose a near-field two dimension (2D) hierarchical beam training scheme. Specifically, codewords are searched in multi-resolution codebooks layer by layer, where angle and distance ranges are reduced gradually. Moreover, we provide the analysis of the proposed beam training overhead, which is proportional to the sum of the logarithm of the antenna number and the sampled distance grid number. Simulation results show that the proposed beam training scheme can reach sub-optimal achievable rate performance with low overhead.

C. Organization and Notations

Organization: The rest of the paper is organized as follows. In Section II, we first introduce the signal model, the near-field channel model, and the formulation of the near-field codebook design problem. In Section III, we provide the design of the theoretical codeword by the proposed Gerchberg-Saxton algorithm considering fully digital architecture. In Section IV, we propose an alternating optimization scheme to design the practical codeword with hybrid digital-analog architecture. Then, the proposed near-field 2D hierarchical beam training scheme is described in Section V. Simulation results and conclusions are provided in Section VI and Section VII, respectively.

Notations: Lower-case and upper-case boldface letters $\mathbf{a}$ and $\mathbf{A}$ denote a vector and a matrix, respectively; $\mathbf{a}^H$ and $\mathbf{A}^H$

denote the conjugate transpose of vector $\mathbf{a}$ and matrix $\mathbf{A}$, respectively; $\|\mathbf{a}\|_2$ denotes the l_2 norm of vector $\mathbf{a}$; $\|\mathbf{a}\|_F$ denotes the Frobenius norm of vector $\mathbf{a}$. $\mathbf{0}_{N\times M}$ denotes $N \times M$-dimensional null matrix. $\mathcal{U}(-a, a)$ denotes the probability density function of uniform distribution on $(-a, a)$. Finally, $\mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes the probability density function of complex multivariate Gaussian distribution with mean $\boldsymbol{\mu}$ and variance $\boldsymbol{\Sigma}$.

II. System Model

In this section, the signal model of the XL-MIMO system is first introduced. Then, the existing near-field XL-MIMO channel model will be briefly reviewed. Finally, we formulate the codeword design problem in the near-field scenario.

A. Signal Model

We consider the XL-MIMO near-field scenario where the BS employs a N-element ELAA serving a single-antenna user. Let $\mathbf{h}^H \in \mathbb{C}^{1\times N}$ denote the channel between the BS and the user. Since the XL-MIMO channel $\mathbf{h}^H$ is generally dominated by a few main paths, we only need to search the physical location of the main paths by beam training instead of acquiring the explicit channel information [17], [18]. Therefore, the main path is concerned in this paper, and the corresponding beam training method will be investigated to search for the optimal beamforming vector to align with the main path.

Take downlink transmission as example, the received signal y can be represented by

$$y = \mathbf{h}^H \mathbf{v} s + n, \tag{1}$$

where $\mathbf{v} \in \mathbb{C}^{N\times 1}$ represents the beamforming vector at the BS, which is essentially a codeword chosen from the predefined codebook, s represents the symbol transmitted by the BS, and $\mathbf{n} \sim \mathcal{CN}\left(\mathbf{0}_N, \sigma^2 \mathbf{I}_N\right)$ represents the received noise with σ^2 representing the noise power. The beam training is measuring the power of the received signal y to find the best codeword from the predefined codebook.

Next, we will briefly review the existing near-field XL-MIMO channel model for existing near-field beam training schemes. It is worth pointing out that the near-field region of wireless communication mentioned in this paper is not the short-range *reactive near-field* communication that happens in the Internet of Things but the radiative near-field communication happens in wireless communication systems. Different from the reactive near-field channel model with noticeable amplitude across the whole antenna array, the amplitude variations across the antennas are negligible in the radiative near-field. If not specifically stated, the "near-field" in this paper means the "radiative near-field".

B. Near-Field XL-MIMO Channel Model

The boundary between the far-field and near-field regions is defined as Rayleigh distance [19], which is calculated by the largest phase discrepancy of $\pi/8$ between planar and spherical wave-
fronts. Specifically, the true phase of the EM wave calculated based on the near-field spherical wave model and far-field planar wave model results in a phase discrepancy. When the largest phase discrepancy among all BS and UE antennas reaches $\pi/8$, the distance between the BS array center and the UE array center is defined as the Rayleigh distance.

When the distance between the BS and the UE is smaller than the Rayleigh distance, the near-field XL-MIMO channel should be modeled under the assumption of spherical wave, which can be expressed by

$$\mathbf{h} = \sqrt{N}\alpha \mathbf{b}\left(\theta, r\right). \tag{2}$$

where α is the complex path gain. Specifically, this path gain α only indicates the small-scale fading. The large-scale fading is assumed to be the same across the array, thus we neglect large-scale fading in the channel model. $\mathbf{b}\left(\theta, r\right)$ represents the near-field array response vector, which can be represented by [11]

$$\mathbf{b}(\theta, r) = \frac{1}{\sqrt{N}}[e^{-j\frac{2\pi}{\lambda}(r^{(1)}-r)}, \cdots, e^{-j\frac{2\pi}{\lambda}(r^{(N)}-r)}]^H, \tag{3}$$

where r represents the distance from the UE to the center of the antenna array,
$r^{(n)} = \sqrt{r^2 + \delta_n^2 d^2 - 2r\delta_n d\theta}$ represents the distance from the UE to the nth BS antenna, and $\delta_n = \frac{2n-N-1}{2}$ with $n = 1, 2, \cdots, N$. It is worth pointing out that a closely spaced array with spacing less than half a wavelength can arise, resulting in interference or coupling between the antennas. In this case, we set the antenna spacing d as half wavelength to avoid coupling between the antennas.

Before data transmission, beam training should be applied to estimate the physical angles and distances of near-field channel paths. it's necessary to conduct beam training to assess the actual angles and distances of the channel paths in the near-field.

The near-field response vector $\mathbf{b}(\theta, r)$ implies that the optimal beam training codeword should focus on the spatial angle θ and BS-UE distance r. The existing near-field beam training scheme is conducting an exhaustive search in the polar-domain codebook [11], which can be represented as

$$\mathbf{A} = [\mathbf{b}(\theta_1, r_1^1), \cdots, \mathbf{b}(\theta_1, r_1^{S_1}), \cdots, \mathbf{b}(\theta_N, r_N^{S_N})], \tag{4}$$

where each column of polar-domain codebook $\mathbf{A}$ is a codeword aligned with the grid $(\theta_n, r_n^{s_n})$, with $s_n = 1, 2, \cdots, S_n$, S_n denotes the number of sampled distance grids at θ_n. Therefore, the number of total sampled grids of the whole propagation environment is $S = \sum_{n=1}^{N} S_n$. Apparently, in XL-MIMO systems, the codebook should not only sample angle but also distance, which leads to a large-size codebook and unfordable beam training overhead. Thus, to address this problem, we design the hierarchical near-field codebook with multi-resolution codebooks, and then propose the corresponding near-field 2D hierarchical beam training. To design the multi-resolution near-field codebooks, we will first formulate the design problem of a near-field codeword with different angle coverage and distance coverage.

C. Formulation of Codebook Design Problem

Suppose the angle coverage and distance coverage of codeword $\mathbf{v}$ are $\mathbf{B}_{\mathbf{v},\theta} \triangleq [\theta, \theta + B_\theta]$ and $\mathbf{B}_{\mathbf{v},r} \triangleq [r, r + B_r]$, where B_θ and B_r are the angle sampled step and distance sampled step. The ideal beam pattern vector of the codeword $\mathbf{v}$ is denote as

$$\mathbf{g}_{\mathbf{v}} = \left[g_{\mathbf{v}}(\theta_1, r_1^1), \cdots, g_{\mathbf{v}}(\theta_N, r_N^1), \cdots, g_{\mathbf{v}}(\theta_N, r_N^{S_N}) \right], \quad (5)$$

where $g_{\mathbf{v}}(\theta, r) = |g_{\mathbf{v}}(\theta, r)|\, e^{j f_{\mathbf{v}}(\theta,r)}$ is the theoretical beamforming gain. The amplitude information $|g_{\mathbf{v}}(\theta, r)|$ of the ideal beam pattern can be further represented by

$$|g_{\mathbf{v}}(\theta, r)| = \begin{cases} \sqrt{C_{\mathbf{v}}}, & \theta \in \mathbf{B}_{\mathbf{v},\theta},\ r \in \mathbf{B}_{\mathbf{v},r} \\ 0, & \theta \notin \mathbf{B}_{\mathbf{v},\theta},\ r \notin \mathbf{B}_{\mathbf{v},r} \end{cases}, \quad (6)$$

where C_v is the absolute value of beamforming gain for the ideal beam pattern. Moreover, the absolute value C_v is related to the power of the codeword [17]. For the ideal beam pattern in (5), the amplitude information $|g_{\mathbf{v}}(\theta, r)|$ of ideal beam pattern vector in target angle coverage and distance coverage are fixed and flattened while other beamforming gains are zero. Meanwhile, the phase information $f_{\mathbf{v}}(\theta, r)$ of the ideal beam pattern vector can be designed flexibly. Compared to a far-field codeword, the near-field codeword should cover not only a certain angle range but also a certain distance range.

To evaluate the effectiveness of the codeword $\mathbf{v}$, we reference $G(\mathbf{v}, \theta, r)$ as the beamforming gain of $\mathbf{v}$ in the angle θ and the distance r. The $G(\mathbf{v}, \theta, r)$ can be represented as

$$G(\mathbf{v}, \theta, r) = \sqrt{N}\mathbf{b}(\theta, r)^H \mathbf{v}. \quad (7)$$

Thus, according to the definition of polar-domain codebook $\mathbf{A}$ in (4), the beam pattern obtained by beamforming with codeword $\mathbf{v}$ can be presented as $\mathbf{A}^H\mathbf{v}$.

The aim of designing a codeword is to make the beam pattern $\mathbf{A}^H\mathbf{v}$ obtained by beamforming with the codeword $\mathbf{v}$ as close as possible to the ideal beam pattern $\mathbf{g}_{\mathbf{v}}$. In order to guarantee the balance between the balance of beamforming gain and location fairness, we formulate the codeword design problem in a minimization formulation [9], [17]. Thus, the objective of the theoretical codeword $\mathbf{v}$ design can be express as

$$\min_{\mathbf{v}, f(\theta,r)} \left\| \mathbf{A}^H\mathbf{v} - \mathbf{g}_{\mathbf{v}} \right\|_2^2. \quad \text{(P1)}$$

In (P1), the ideal theoretical codeword $\mathbf{v}$ can only be realized by the fully digital architecture, where each antenna requires one dedicated radio frequency (RF) chain to realize fully digital signal processing. However, fully digital architecture in the XL-MIMO system results in unaffordable energy consumption. In fact, a hybrid digital-analog structure is usually preferred in XL-MIMO systems to improve energy efficiency [20]. In this structure, we need to design practical codewords considering the hardware constraints of the phase shifter resolution and the number of radio frequency (RF) chains N_{RF} [21].

Specifically, based on the ideal theoretical codeword $\mathbf{v}$, the design of the practical codeword $\mathbf{v}_{\mathrm{p}} \triangleq \mathbf{F}_{\mathrm{RF}}\mathbf{f}_{\mathrm{BB}}$ can be represented as

$$\begin{aligned} \min_{\mathbf{F}_{\mathrm{RF}}, \mathbf{f}_{\mathrm{BB}}} \ & \|\mathbf{v} - \mathbf{F}_{\mathrm{RF}}\mathbf{f}_{\mathrm{BB}}\|_2 \\ \text{s.t.} \ & \|\mathbf{F}_{\mathrm{RF}}\mathbf{f}_{\mathrm{BB}}\|_2 = 1, \\ & [\mathbf{F}_{\mathrm{RF}}]_{n,i} = e^{j\delta_{n,i}}, \delta_{n,i} \in \Phi_b \\ & n = 1, 2, \ldots, N, i = 1, 2, \ldots, N_{\mathrm{RF}}, \end{aligned} \quad \text{(P2)}$$

where the $\mathbf{F}_{\mathrm{RF}} \in \mathbb{C}^{N \times N_{\mathrm{RF}}}$ and $\mathbf{f}_{\mathrm{BB}} \in \mathbb{C}^{N_{\mathrm{RF}} \times 1}$ are the analog beamforming matrix and the digital beamforming vector. $\mathbf{\Phi}_b = \left[\pi\left(-1 + \frac{1}{2^b}\right), \pi\left(-1 + \frac{3}{2^b}\right), \ldots \pi\left(1 - \frac{1}{2^b}\right)\right]$ is the set of quantized phase shifters with b bits.

All the codewords in the codebook can be designed based on (P1) and (P2). Next, we introduce the design method of the theoretical codeword $\mathbf{v}$ in Section III and practical codeword $\mathbf{v}_{\mathrm{p}}$ Section IV.

III. Proposed Gerchberg-Saxton Algorithm based Near-Field Theoretical Codeword Design

In this section, we will first briefly review the Gerchberg-Saxton algorithm applied in the phase retrial problem in the hologram optical system, and the relationship between the phase retrieval problem and the codeword design problem is analyzed. Next, we propose a GS-based theoretical codeword design scheme. Finally, the convergence property of the GS algorithm in near-field codeword design is provided.

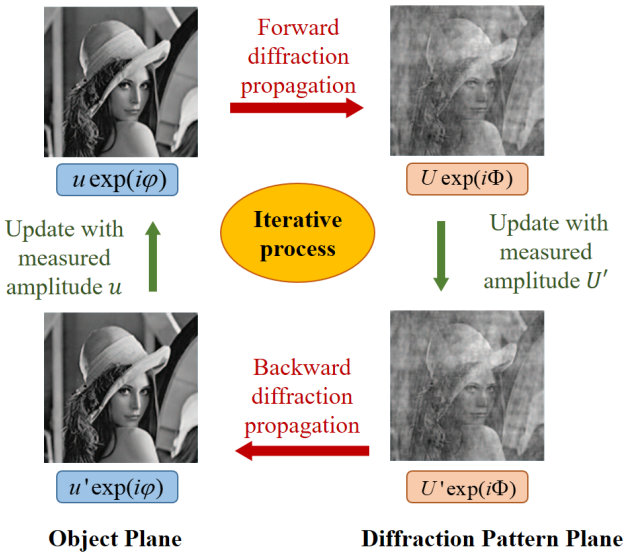
A. Preview of the Phase Retrieval Problem and Gerchberg-Saxton algorithm

1) Phase retrieval problem in digital holography imaging: In recent years, with the development of modern optics and computer science, digital holography imaging technology has changed the traditional imaging object-image relationship and structure by combining the front-end optical system design with the back-end signal processing. The back-end signal processing algorithm of the original data collected by the camera can break through the traditional imaging bottleneck.

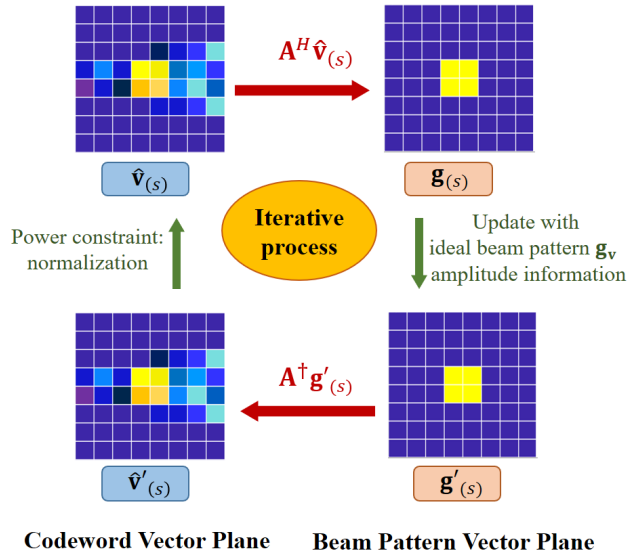
In specific, in optical systems, the amplitude information is easy to measure, while the direct recording of the phase information is not allowed. The reason is that the electromagnetic field oscillates at a very high frequency that rare electronic measurement devices can follow [22]. Thus, in order to realize the imaging of the original object, one of the most important problems in digital holography imaging technology is conducting phase retrieval [23]. Fortunately, with the help of the measured amplitude information, some signal processing algorithms offer alternative methods for recovering the phase information of optical images without requiring sophisticated devices.

Reviewing the theoretical codeword design problem in (P1), it is obvious that the problem (P1) is similar to the phase retrieval in digital holography imaging, where the phase information ($f_{\mathbf{v}}(\theta, r)$ of the ideal beam pattern vector) should be obtained by measured amplitude information ($|g_{\mathbf{v}}(\theta, r)|$ of ideal beam pattern vector).

This article has been accepted for publication in IEEE Transactions on Communications. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TCOMM.2023.3344600



(a) Illustration of GS algorithm in iterative phase retrieval problem.



(b) Illustration of GS algorithm in codeword design

Fig. 1. Comparisons of the original and improved GS algorithm.

2) Gerchberg-Saxton algorithm: One of the most popular methods to solve the phase retrieval problem is Gerchberg–Saxton (GS)-based algorithm [24], [25] as shown in Fig. 1 (a), where two amplitude measurements are iteratively imposed in the object plane and diffraction pattern plane [26]. It is worth noting that the diffraction pattern plane is also known as the Fourier plane since the complex-valued wavefronts in the object and the diffraction pattern planes are usually connected through a Fourier transform with each other.

Specifically, the GS algorithm initializes in the object plane, where the initial complex-valued wavefronts are created by combining the measured amplitude information with the random phase information. The iteration process of the GS algorithm consists of four steps: i) The forward diffraction propagation of the wavefronts in the object plane provides complex-valued wavefronts in the diffraction pattern plane; ii) Update the complex-valued wavefronts in the diffraction pattern plane: the amplitude information is substituted with the measured amplitude information U'; iii) The backward diffraction propagation provides the complex-valued wavefronts in the object plane; iv) Update the complex-valued wavefronts in the diffraction plane: The amplitude information in the object plane is substituted with the measured amplitude information. The result of the GS algorithm is the recovered complex-valued wavefronts in the diffraction pattern plane.

Some modified versions of the GS algorithm have been proposed afterward [27] to match various imaging problems. Instead of utilizing the GS algorithm in the imaging problem, we improved the GS algorithm in the near-field codeword design problem. In specific, we replace one of updating processes with measured amplitude information by applying normalization to match the power constraint of the codeword.

B. Design of the Theoretical Codeword $\mathbf{v}$

In order to solve the (P1), we draw the experience from the Gerchberg–Saxton (GS) algorithm, which is widely applied in phase retrieval problems in digital hologram imaging of optical systems. In the phase retrieval problem, the phase information needed to be obtained with the fixed amplitude information, which is the same as the phase information $f_\mathbf{v}(\theta, r)$ design of the ideal beam pattern in the problem (P1). Specifically, the proposed GS-based near-field codeword design procedure is shown in **Algorithm 1**.

Algorithm 1: GS-based theoretical codeword design

Inputs: $|\mathbf{g}_\mathbf{v}|$, $C_\mathbf{v}$, S_{max}, $\mathbf{A}$, $B_{\mathbf{v},\theta}$, $B_{\mathbf{v},r}$.
Initialization: randomly generate $f_{(0)}(\theta, r)$ and obtain the $\mathbf{g}_{(0)}$ by (8).
1. $\hat{\mathbf{v}}'_{(0)} = \left(\mathbf{A}\mathbf{A}^H\right)^{-1}\mathbf{A}\mathbf{g}_{(0)}$
2. Obtain $\hat{\mathbf{v}}_{(1)}$ by normalizing $\hat{\mathbf{v}}'_{(0)}$
3. **for** $s = 1, 2, \cdots, S_{\max}$ **do**
4. calculate $\mathbf{g}_{(s)}$ based on $\hat{\mathbf{v}}_{(s)}$ by (9)
5. calculate $\mathbf{g}'_{(s)}$ based on $\mathbf{g}_{(s)}$ and $\mathbf{g}_\mathbf{v}$ by (10)
6. calculate $\hat{\mathbf{v}}'_{(s)}$ based on $\mathbf{g}'_{(s)}$ by (11)
7. **if** $s < S_{max}$
8. calculate $\hat{\mathbf{v}}_{(s+1)}$ based on $\hat{\mathbf{v}}'_{(s)}$ by (12)
9. **end if**
10. **end for**
11. $\mathbf{v} = \hat{\mathbf{v}}'_{(S_{max})}/||\hat{\mathbf{v}}'_{(S_{max})}||_2$
Output: Theoretical codeword $\mathbf{v}$.

For notation simplicity, in the description of the GS algorithm, we use $\hat{\mathbf{v}}_{(s)}$, $\mathbf{g}_{(s)}$, $\mathbf{g}'_{(s)}$, and $\hat{\mathbf{v}}'_{(s)}$ to denote the designed codeword vector, the beam pattern vector realized by the designed codeword, the revised beam pattern vector with ideal beam pattern amplitude, and the codeword vector obtained by revised beam pattern vector in the s-th iteration of GS algorithm.

Before the GS algorithm starts, we should first obtain the initial beam pattern vector $\mathbf{g}_{(0)}$ with randomly generated phase

$f_{(0)}(\theta, r)$ and amplitude information $g_{\mathbf{v}}(\theta, r)$ of ideal beam pattern vector $\mathbf{g}_{\mathbf{v}}$. In this way, the $\mathbf{g}_{(0)}$ can be represented as

$$\mathbf{g}_{(0)}=\Big[\,|g_{\mathbf{v}}(\theta_1, r_1^1)|f_{(0)}(\theta_1, r_1^1),\cdots,|g_{\mathbf{v}}(\theta_N, r_N^1)|f_{(0)}(\theta_N, r_N^1), \cdots, |g_{\mathbf{v}}(\theta_N, r_N^{S_N})|f_{(0)}(\theta_N, r_N^{S_N})\Big]. \tag{8}$$

In s-th iteration, with provided designed $\hat{\mathbf{v}}_{(s)}$,

$$\mathbf{g}_{(s)} = \mathbf{A}^H \hat{\mathbf{v}}_{(s)}. \tag{9}$$

Then, in order to maintain the amplitude information of the ideal beam pattern vector $\mathbf{g}_{\mathbf{v}}$ to approach the ideal beam pattern, we assign the amplitude information $|\mathbf{g}_{\mathbf{v}}(\theta, r)|$ of ideal beam pattern $\mathbf{g}_{\mathbf{v}}$ to $\mathbf{g'}_{(s)}$, and the phase information $f_{\hat{\mathbf{v}}_{(s)}}(\theta, r)$ of current beam pattern $\mathbf{g}_{(s)}$ to $\mathbf{g'}_{(s)}$. In this case, the $\mathbf{g'}_{(s)}$ can be presented as

$$\mathbf{g'}_{(s)}=\Big[\,|g_{\mathbf{v}}(\theta_1, r_1^1)|f_{\hat{\mathbf{v}}_{(s)}}(\theta_1, r_1^1), \cdots, |g_{\mathbf{v}}(\theta_N, r_N^1)|f_{\hat{\mathbf{v}}_{(s)}}(\theta_N, r_N^1), \cdots, |g_{\mathbf{v}}(\theta_N, r_N^{S_N})|f_{\hat{\mathbf{v}}_{(s)}}(\theta_N, r_N^{S_N})\Big]. \tag{10}$$

Base on the (P1), given $\mathbf{g'}_{(s)}$, the $\hat{\mathbf{v}}'_{(s)}$ can be obtained by least square algorithm as

$$\hat{\mathbf{v}}'_{(s)} = \left(\mathbf{A}\mathbf{A}^H\right)^{-1}\mathbf{A}\mathbf{g}'_{(s)} = \mathbf{A}^{\dagger}\mathbf{g}'_{(s)}, \tag{11}$$

where the pseudo inverse of $\mathbf{A}^H$ is denoted as $\mathbf{A}^{\dagger}$. Finally, we normalize the $\hat{\mathbf{v}}'_{(s)}$ as

$$\hat{\mathbf{v}}_{(s+1)} = \hat{\mathbf{v}}'_{(s)}/||\hat{\mathbf{v}}'_{(s)}||_2. \tag{12}$$

After the iteration number reaches S_{max}, we utilize $\hat{\mathbf{v}}'_{(S_{max})}$ to obtain the designed theoretical codeword $\mathbf{v}$.

From the description of the **Algorithm 1** above, we can see that Step 6 involves the beamforming gain C_v. Importantly, this gain C_v is not constrained by power limitation. We normalize the ideal codeword to satisfy the power constraint $||\mathbf{v}||^2 = 1$ at the end of the **Algorithm 1**. Thus, in our work, we operate under the assumption that the input beamforming gain $C_v = 1$ as indicated in [17], where we temporarily omit the power constraint.

C. Convergence Property of GS Algorithm in Near-Field Codeword Design

As mentioned before, the original GS algorithm assumes that the object and the diffraction pattern planes are connected through a Fourier Transform (FT). The convergence of the original GS algorithm with FT assumption is proved based on Parseval's theorem of FT [28], where the energy of wavefronts in the object and the diffraction pattern planes before and after FT and inverse FT are the same. However, the codeword vector plane and beam pattern vector plane in the proposed GS algorithm are connected with the polar-domain transformation, which does not satisfy Parseval's theorem. Thus, the convergence property of the proposed GS algorithm based on polar-domain transformation in near-field codeword design should be analyzed.

In this paper, the convergence of the proposed GS algorithm is supervised by the squared error in each iteration. Specifically, the squared error of the beam pattern plane in s-th iteration can be presented as

$$\begin{aligned} E_{(s)} &= \iint \|\mathbf{g}_{(s)}(\theta, r) - \mathbf{g'}_{(s)}(\theta, r)\|_2^2 d\theta dr. \\ &= \iint \|\mathbf{A}^H\hat{\mathbf{v}}_{(s)}(u, w) - \mathbf{A}^H\hat{\mathbf{v}}'_{(s)}(u, w)\|_2^2 du dw \end{aligned} \tag{13}$$

It is worth noting that the codewords in the polar-domain codebook $\mathbf{A}^H$ have been rearranged, where the codewords aligned with the largest distance S_n of each θ_n are brought to the front columns of $\mathbf{A}^H$. Thus, the $\mathbf{A}^H$ can be rewritten as

$$\mathbf{A}^H = [\mathbf{A}_1, \mathbf{A}_2]^H, \tag{14}$$

where $\mathbf{A}_1 = [\mathbf{b}(\theta_1, r_1^{S_1}), \mathbf{b}(\theta_2, r_2^{S_2}), \cdots, \mathbf{b}(\theta_N, r_N^{S_N})]$,

$\mathbf{A}_2 = [\mathbf{b}(\theta_1, r_1^1), \cdots, \mathbf{b}(\theta_1, r_1^{S_1-1}), \cdots, \mathbf{b}(\theta_N, r_1^1), \cdots, \mathbf{b}(\theta_N, r_N^{S_N-1})]$.

Since the S_n in each column $\mathbf{b}(\theta_n, r_1^{S_n})$ of $\mathbf{A}_1$ is larger than Rayleigh distance, $\mathbf{b}(\theta_n, r_1^{S_n})$ approximates to the far-field codeword aligned with the physical direction θ_n. In this case, the $\mathbf{A}_1$ is equal to a far-field DFT codebook. Thus, $\mathbf{A}_1^H\left(\hat{\mathbf{v}}_{(s)}(u, w) - \hat{\mathbf{v}}'_{(s)}(u, w)\right)$ is an FT process, which satisfies Parseval's theorem as

$$\begin{aligned} &\iint \| \left(\hat{\mathbf{v}}_{(s)}(u, w) - \hat{\mathbf{v}}'_{(s)}(u, w)\right) \|_2^2 du dw \\ =&\iint \|\mathbf{A}_1^H \left(\hat{\mathbf{v}}_{(s)}(u, w) - \hat{\mathbf{v}}'_{(s)}(u, w)\right) \|_2^2 du dw. \end{aligned} \tag{15}$$

Therefore, $E_{(s)}$ can be further expressed as

$$\begin{aligned} E_{(s)} =&\iint \|\mathbf{A}_1^H \left(\hat{\mathbf{v}}_{(s)}(u, w) - \hat{\mathbf{v}}'_{(s)}(u, w)\right) \|_2^2 \\ &+ \|\mathbf{A}_2^H \left(\hat{\mathbf{v}}_{(s)}(u, w) - \hat{\mathbf{v}}'_{(s)}(u, w)\right) \|_2^2 du dw. \\ \geq&\iint \| \left(\hat{\mathbf{v}}_{(s)}(u, w) - \hat{\mathbf{v}}'_{(s)}(u, w)\right) \|_2^2 \end{aligned} \tag{16}$$

The squared error of the codeword vector plane of $s+1$-th iteration for the GS algorithm can be expressed as

$$E^0_{(s)} = \iint \|\hat{\mathbf{v}}_{(s+1)}(u, w) - \hat{\mathbf{v}}'_{(s)}(u, w)\|_2^2 du dw. \tag{17}$$

Then, we provide **Lemma 1** to show the change of squared error between adjacent iterations in the codeword vector plane.

Lemma 1: *In the codeword vector plane of GS algorithm, the error between* $\hat{\mathbf{v}}_{(s)}(u, w)$ *and* $\hat{\mathbf{v}}'_{(s)}(u, w)$ *not less than than the error between* $\hat{\mathbf{v}}_{(s+1)}(u, w)$ *and* $\hat{\mathbf{v}}'_{(s)}(u, w)$*, i.e.,* $\|\hat{\mathbf{v}}_{(s)}(u, w) - \hat{\mathbf{v}}'_{(s)}(u, w)\|_2^2 > \|\hat{\mathbf{v}}_{(s+1)}(u, w) - \hat{\mathbf{v}}'_{(s)}(u, w)\|_2^2$.

proof: See Appendix A.

From the (16), (17), and **Lemma 1**, we can derive that

$$E^0_{(s)} \leq \iint \|\hat{\mathbf{v}}_{(s)}(u, w) - \hat{\mathbf{v}}'_{(s)}(u, w)\|_2^2 du dw \leq E_{(s)} \tag{18}$$

On the other hand, $E^0_{(s)}$ can be further expressed as

$$\begin{aligned} E^0_{(s)} &= \iint \|\hat{\mathbf{v}}_{(s+1)}(u, w) - \hat{\mathbf{v}}'_{(s)}(u, w)\|_2^2 du dw \\ &= \iint \|\mathbf{A}^{\dagger}\mathbf{g}_{(s+1)}(\theta, r) - \mathbf{A}^{\dagger}\mathbf{g}'_{(s)}(\theta, r)\|_2^2 d\theta dr. \end{aligned} \tag{19}$$

It is easy to obtain that

$$\begin{aligned}\mathbf{A}\left[(\mathbf{A}_1^H)^{-1},\mathbf{0}_{(S-N)\times N}\right] &= [\mathbf{A}_1,\mathbf{A}_2]\left[(\mathbf{A}_1^H)^{-1},\mathbf{0}_{(S-N)\times N}\right]^H \\ &= \mathbf{I}_{N\times N}.\end{aligned} \tag{20}$$

In this case, utilizing the uniqueness of pseudo inverses [29], we can easily know that pseudo inverses of matrix $\mathbf{A}$ can be denoted as $\mathbf{A}^\dagger = \left[(\mathbf{A}_1^H)^{-1},\mathbf{0}_{(S-N)\times N}\right]$. Then, we can substitute the $\mathbf{A}^\dagger$ into (19) by $\left[(\mathbf{A}_1^H)^{-1},\mathbf{0}_{(S-N)\times N}\right]$, and the (19) can be further expressed as

$$\begin{aligned}E_{(s)}^0 &= \iint \|\mathbf{A}_1^H\mathbf{g}_{(s+1)}(\theta,r) - \mathbf{A}_1^H\mathbf{g}'_{(s)}(\theta,r)\|_2^2 d\theta dr \\ &= \iint \|\mathbf{A}_1^H\left(\mathbf{g}_{(s+1)}(\theta,r) - \mathbf{g}'_{(s)}(\theta,r)\right)\|_2^2 d\theta dr\end{aligned}. \tag{21}$$

Since $\mathbf{A}_1^H$ is also a orthogonal DFT codebook, $\mathbf{A}_1^H\left(\mathbf{g}_{(s+1)}(\theta,r) - \mathbf{g}'_{(s)}(\theta,r)\right)$ in (21) is an inverse FT process. Thus, the (21) also satisfies Parseval's theorem as (15), which can be further expressed as

$$E_{(s)}^0 = \iint \|\mathbf{g}_{(s+1)}(\theta,r) - \mathbf{g}'_{(s)}(\theta,r)\|_2^2 d\theta dr. \tag{22}$$

Similar to **Lemma 1**, we can obtain that

$$\|\mathbf{g}_{(s+1)}(\theta,r) - \mathbf{g}'_{(s)}(\theta,r)\|_2^2 \geq \|\mathbf{g}_{(s+1)}(\theta,r) - \mathbf{g}'_{(s+1)}(\theta,r)\|_2^2 \tag{23}$$

Thus, given (22) and (23)

$$E_{(s)}^0 \geq \iint \|\mathbf{g}_{(s+1)}(\theta,r) - \mathbf{g}'_{(s+1)}(\theta,r)\|_2^2 d\theta dr = E_{(s+1)}. \tag{24}$$

Combining the equation (18) and (24), we can observe that

$$E_{(s+1)} \leq E_{(s)}^0 \leq E_{(s)}, \tag{25}$$

which means that the squared error in each iteration decreases. Thus, the convergence property of the proposed GS algorithm is proven.

IV. Proposed Alternating Optimization based Near-Field Practical Codeword Design

It is well known that each antenna requires one dedicated radio-frequency (RF) chain to realize the fully digital architecture. In this way, an XL-MIMO system with a very large number of antennas leads to an equally large number of RF chains, which will result in unaffordable hardware costs and energy consumption. To solve this problem, hybrid digital-analog architecture is preferred in practice, where the fully digital beamforming matrix is decomposed into a high-dimensional analog beamforming matrix and a low-dimensional digital beamforming vector. Moreover, quantized phase shifts instead of continuous quantized phase shifts are accessible for realizing analog beamforming matrix. Thus, in this section, alternating optimization is proposed for practical codeword design considering the hybrid digital-analog architecture and quantized phase shifts.

Based on the theoretical codeword $\mathbf{v}$ obtained by **Algorithm 1**, we solve the practical codeword $\mathbf{v}_\mathrm{p}$ design problem (P2) by alternating optimizing the digital beamforming vector $\mathbf{f}_\mathrm{BB}$ and the analog beamforming matrix $\mathbf{F}_\mathrm{RF}$ considering the hardware constraints. **Algorithm 2** provides the specific procedure to design the practical codeword.

Algorithm 2: Practical codeword design

Inputs: $\mathbf{v}$, $\mathbf{T}_{max}$, $\mathbf{P}_{max}$, $\mathbf{\Phi}_b$, N, N_RF.
Initialization: randomly generate $\mathbf{F}_\mathrm{RF}^0$.
1. **for** $t = 1, 2, \cdots, T_\mathrm{max}$ **do**
// Design the digital beamforming vector.
2. calculate the $\mathbf{f}_\mathrm{BB}^t$ by (26)
// Design the analog beamforming matrix.
3. **for** $p = 1, 2, \cdots, P_{max}$ **do**
4. **for** $n = 1, 2, \cdots, N$ **do**
5. **for** $i = 1, 2, \cdots, N_\mathrm{RF}$ **do**
6. Search $\delta_{n,i}$ to satisfy (27)
7. **end for**
8. **end for**
9. **if** $\delta_{n,i}^{p-1} = \delta_{n,i}^p$ **then**
10. Jump to Step2
11. **end if**
12. **end for**
13. obtain the $\mathbf{F}_\mathrm{RF}^t$ by utilizing (28)
14. **end for**
Output: $\mathbf{f}_\mathrm{BB} = \mathbf{f}_\mathrm{BB}^{T_{max}}$, $\mathbf{F}_\mathrm{RF} = \mathbf{F}_\mathrm{RF}^{T_{max}}$, $\mathbf{v}_p = \mathbf{F}_\mathrm{RF}^{T_{max}}\mathbf{f}_\mathrm{BB}^{T_{max}}$

For the given analog beamforming matrix $\mathbf{F}_\mathrm{RF}$, the optimization problem of the digital beamforming vector $\mathbf{f}_\mathrm{BB}$ can be expressed as

$$\min_{\mathbf{f}_\mathrm{BB}} \|\mathbf{v} - \mathbf{F}_\mathrm{RF}\mathbf{f}_\mathrm{BB}\|_2, \tag{P2.1}$$

which can be solved by least square as

$$\hat{\mathbf{f}}_\mathrm{BB} = \left(\mathbf{F}_\mathrm{RF}^H\mathbf{F}_\mathrm{RF}\right)^{-1}\mathbf{F}_\mathrm{RF}^H\mathbf{v} \tag{26}$$

Then, for the given analog beamforming vector $\mathbf{f}_\mathrm{BB}$, the optimization problem of $\mathbf{F}_\mathrm{RF}$ can be expressed as

$$\begin{aligned}\min_{\mathbf{F}_\mathrm{RF}} \ & \|\mathbf{v} - \mathbf{F}_\mathrm{RF}\mathbf{f}_\mathrm{BB}\|_2 \\ \text{s.t.} \ & \|\mathbf{F}_\mathrm{RF}\mathbf{f}_\mathrm{BB}\|_2 = 1, \\ & [\mathbf{F}_\mathrm{RF}]_{n,i} = e^{j\delta_{n,i}}, \delta_{n,i} \in \Phi_b, \\ & n = 1, 2, \ldots, N, i = 1, 2, \ldots, N_\mathrm{RF},\end{aligned} \tag{P2.2}$$

The optimization of $\mathbf{F}_\mathrm{RF}$ problem (P2.2) can be converted to the minimization absolute value of each entry of the vector $\mathbf{v} - \mathbf{F}_\mathrm{RF}\mathbf{f}_\mathrm{BB}$. Hence, the problem (P2.2) can be transformed into N sub-problems, which can be optimized one by one. The n-th sub-problem is rewritten as

$$\begin{aligned}\min_{\theta_1,\theta_2,\ldots,\theta_{N_\mathrm{RF}}} & \left|[\mathbf{v}]_n - \sum_{i=1}^{N_\mathrm{RF}}[\mathbf{f}_\mathrm{BB}]_i\, e^{j\delta_{n,i}}\right| \\ \text{s.t. } & \delta_{n,i} \in \Phi_b, i = 1, 2, \ldots, N_\mathrm{RF}.\end{aligned} \tag{27}$$

To obtain the solution to (27), the exhaustive search is a obvious choice, where all the combination of $\delta_{n,1}, \cdots, \delta_{n,N_\mathrm{RF}}$ are test to minimize the objective. However, the number of combination is 2^{bN_RF}, which has prohibitively high computational complexity. For example, if $b = 4, N_\mathrm{RF} = 32$, the

$2^{bN_{\mathrm{RF}}} \approx 7.9\times10^{28}$! Thus, we need to investigate near-optimal search method to reduce complexity.

In this case, we propose a high efficient individual search method, where each $\delta_{n,i}$ is determined separately in each iteration. The specific procedures are summarized in **Algorithm 2**. We firstly initialize the $\delta^0_{n,1},\cdots,\delta^0_{n,N_{\mathrm{RF}}}$ by choosing the entry from the $\mathbf{\Phi}_b$ and generate $\mathbf{F}^0_{\mathrm{RF}}$. In p-th iteration, we find best $\delta_{n,1},\cdots,\delta_{n,N_{\mathrm{RF}}}$ one by one. In step 6, for $\delta_{n,i}$, we search through the $\mathbf{\Phi}_b$ to find the optimal choice to satisfy the (27), i.e., minimizing the loss between the theoretical codeword and the practical codeword. Therefore, in each iteration, the **Algorithm 2** naturally satisfies the convergence condition that the square error between the theoretical codeword and the practical codeword in each iteration decreases. This iterative process performs stop until the number of iterations reaches predetermined figure or $\delta^{p-1}_{n,i}=\delta^p_{n,i}$. Then the n-th row of the designed $\hat{\mathbf{F}}_{\mathrm{RF}}$ can be expressed as

$$\left[\hat{\mathbf{F}}_{\mathrm{RF}}\right]_{n,:}=\left[e^{j\hat{\delta}_{n,1}},e^{j\hat{\delta}_{n,2}},\ldots,e^{j\hat{\delta}_{n,N_{\mathrm{RF}}}}\right] \tag{28}$$

After T_{max} iteration, we can obtain the final practical codeword as

$$\mathbf{v}_p=\mathbf{F}^{T_{max}}_{\mathrm{RF}}\mathbf{f}^{T_{max}}_{\mathrm{BB}} \tag{29}$$

The proposed practical codeword design algorithm can be extend to fully analog structure, where the antennas on the array are all connected to one RF chain. In specific, after the ideal theoretical codeword is obtained by the **Algorithm 1**, instead of iteratively designing the digital beamforming vector and the analog beamforming matrix, we only need to design the analog beamforming matrix by the high efficient individual search method shown in Step 3-13 of **Algorithm 2**.

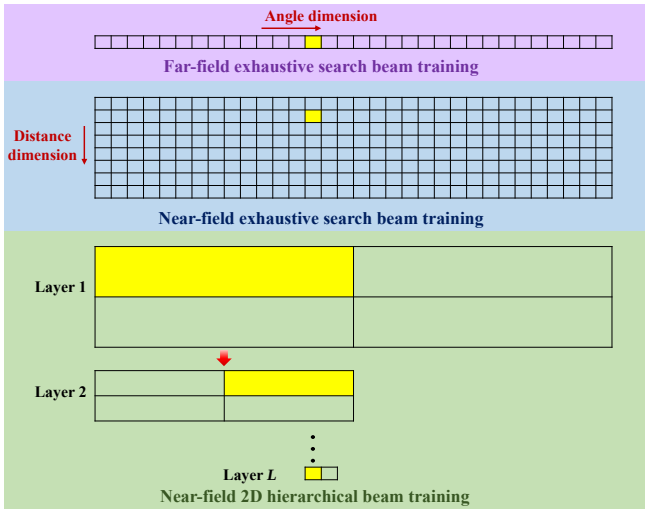


Fig. 2. Comparison between the far-field exhaustive search, near-field exhaustive search and the near-field 2D hierarchical beam training.

It is worth noting that, the proposed codeword design scheme is suitable for not only single-path scenarios but also multi-path scenarios. For a single-path scenario, the beam formed by the chosen codewords can aligned to the physical location of the user. For the multi-path scenario, all the paths will combined into a *equivalent channel*, which means that the user lies in a *equivalent location*. In this case, the beam training scheme will choose the codeword from the predefined codebooks that aligned to the corresponding *equivalent location* to maximize the received signal power. Intuitively, when the number of channel paths is reduced to only one, the multi-path channel is degraded into a single path situation.

V. Proposed Near-Field 2D Hierarchical Beam Training

In this section, we first introduce the proposed near-field 2D hierarchical beam training scheme, where the angle and distance ranges are reduced gradually layer by layer in multi-resolution codebooks. Then, the analysis of the proposed beam training overhead is provided.

A. Near-Field 2D Hierarchical Beam Training Scheme

In order to obtain the tradeoff between the near-field beam training overhead and the performance, one of the methods is to apply a hierarchical near-field codebook, which consists of multi-resolution codebooks. The sizes of codebooks are determined by the angle sample step and distance sample step of the codebook, i.e., B_θ and B_r in (6). Specifically, as the increase of B_θ and B_r, the corresponding codeword has a lower resolution, and the size of the corresponding codebook becomes smaller. As mentioned before, we can generate near-field multi-resolution codebooks with different angle coverage and distance coverage based on the **Algorithm 1** and **Algorithm 2**.

Then, these multi-resolution codebooks are applied to conduct near-field 2D hierarchical beam training. Compared with far-field scenario, the near-field 2D hierarchical beam training need to reduce the search range of angle and distance at the same time as shown in Fig. 2.

The specific near-field beam training procedure is summarized in **Algorithm 3**. First, as shown in Step2, for l-th codebook generation, we need to divide the angle coverage $\mathbf{B}^l_{\mathbf{v}_k,\theta}$ and distance coverage $\mathbf{B}^l_{\mathbf{v}_k,r}$ based on angle samples step B^l_θ and distance samples step B^l_r for each codeword $\mathbf{v}_k$. Then, in Steps 3-4, the codewords design scheme based on **Algorithm 1** and **Algorithm 2** is applied to obtain the l-th codebook $\mathbf{W}_l$. Then, Steps 7-16 are operated to search the optimal codeword in multi-resolution codebooks layer by layer.

B. Comparison of the Beam Training Overhead

Beam training overhead refers to the number of time slots used for beam training. Generally, the beam training overhead is determined by the spatial resolutions of an antenna array on the angle and distance, i.e., the number of sampled angle grids U and the number of sampled distance grids S. It is worth pointing out that U is usually set as the same as the number of antennas on the array. The training overhead of the exhaustive near-field beam training scheme is US. Meanwhile, the training overhead of the time-delay based beam training is only related to the number of sampled distance grids S. For the proposed 2D hierarchical beam training method, the beam

Algorithm 3: Near-field 2D hierarchical beam training

Inputs: L,$\left\{B_\theta^1, B_\theta^2, \cdots, B_\theta^L\right\}$, $\left\{B_r^1, B_r^2, \cdots, B_r^L\right\}$, $y_{opt}=0$, $s_{opt}=0$

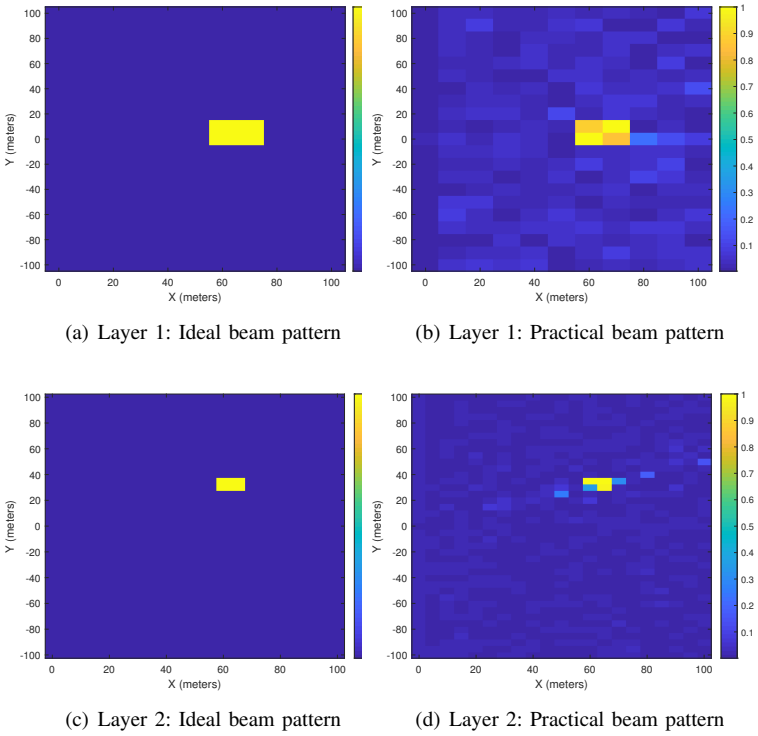
// Generate L sub-codebooks

1. **for** $l=1,2,\cdots,L$ **do**
2. generate the collection of $\mathbf{B}^l_{\mathbf{v}_{l,k},\theta}$ and $\mathbf{B}^l_{\mathbf{v}_{l,k},r}$ based on B_θ^l and B_r^l
3. generate $|g_{\mathbf{v}}(\theta,r)|$ for based on (6)
4. obtain the practical codewords in l-th sub-codebook $\mathbf{W}_l$ based on **Algorithm 1** and **Algorithm 2**.
5. **end for**
6. $\mathbf{W}=\mathbf{W}_1$

// Conduct beam training

7. **for** $l=1,2,\cdots,L$ **do**
8. **for** $\mathbf{v}_{l,k}$ in $\mathbf{W}$ **do**
9. $y_k^l=\mathbf{h}^H\mathbf{v}_{l,k}s+n$
10. **if** $y_k^l>y_{opt}$ **then**
11. $k_{opt}=k$
12. **end if**
13. **end for**
14. choose $\mathbf{v}_{l+1,k}$ in $\mathbf{W}_{l+1}$ satisfied $\mathbf{B}^{l+1}_{\mathbf{v}_{l+1,k},\theta}\in\mathbf{B}^l_{\mathbf{v}_{l,k_{opt}},\theta}$ and $\mathbf{B}^{l+1}_{\mathbf{v}_{l+1,k},r}\in\mathbf{B}^l_{\mathbf{v}_{l,k_{opt}},r}$
15. the chosen codewords $\mathbf{v}_{l+1,k}$ compose the $\mathbf{W}$
16. **end for**

Output: The feedback optimal codeword index k_{opt} from the user.



(a) Layer 1: Ideal beam pattern (b) Layer 1: Practical beam pattern

(c) Layer 2: Ideal beam pattern (d) Layer 2: Practical beam pattern

Fig. 3. Comparison of the beam patterns of different layers of the hierarchical codebook.

training overhead can be represented as $\mathcal{O}\left(log\left(U\right)+log\left(S\right)\right)$. It is obvious that, the training overhead of the proposed 2D hierarchical beam training is much less than that of the exhaustive near-field beam training. Since the number of sampled angle grids U is usually large than the number of sampled distances S [15], the training overhead of the proposed 2D hierarchical beam training is larger than that of the time-delay based beam training. However, the performance of the time-delay based beam training heavily depends on the extra hardware overhead and wideband condition, which will be further verified by simulation results in Section VI.

VI. Simulation Results

For simulations, we assume that the number of BS antennas and RF chains are $N=512$ and $N_{\mathrm{RF}}=100$ [30]. The wavelength is set as $\lambda=0.005$ meters, corresponding to the 60 GHz frequency [30]. Considering the antenna spacing is half-wavelength, the array aperture at BS can be calculated as $N\times\frac{\lambda}{2}=512\times\frac{0.005}{2}=1.28$ m. According to the definition of Rayleigh distance (RD), i.e., $\mathrm{RD}=\frac{2D^2}{\lambda}$, the boundary of near-field and far-field region of our simulation set up is $\mathrm{RD}=\frac{2D^2}{\lambda}=\frac{2\times1.28^2}{0.005}=655$ m. The distance r is generated as $r_l\sim\mathcal{U}\left(20,100\right)$ meters, which is smaller than the RD. Thus, the corresponding channel is in near-field region. The quantified bits number of phase shifters is set as $b=5$. The path gain α caused by small-scale fading, and angle θ are generated as following: $\alpha_l\sim\mathcal{CN}\left(0,1\right)$, and $\theta_l\sim\mathcal{U}\left(-1,1\right)$ [11], [13]. The SNR is defined as $1/\sigma^2$.

Fig. 3 shows the comparison of the ideal beam pattern and the normalized practical beam pattern obtained by conducting beamforming with the designed codeword. In these heat maps, the brighter the color, the greater the beamforming gain at this position. It is worth noting that, we utilize the rectangular coordinate system to present the beamforming gains of the locations in two-dimension space to show the beam pattern more clearly, where the coordinates of the X-axis and Y-axis satisfy $x=r\cos(\theta)$, and $y=r\cos(\theta)$. Fig. 3 (a) presents an ideal beam pattern of the layer 1 codebook, where the beam should focus on the target location, i.e., $x=[55,75], y=[-5,15]$. After we conduct beamforming with the designed practical codeword, we can obtain Fig. 3 (b), which presents the beamforming gains of different locations in space with the designed practical codeword. From Fig. 3 (b) we can see that the target location has the largest beamforming gain and other locations have much lower beamforming gains. Moreover, for the codeword in the layer 2 codebook, the designed practical codeword can also approach the ideal beam pattern Fig. 3 (c) and (d). Since the codeword in the layer 1 codebook should cover a larger range than that of layer 2 codebook, we can observe that the beamforming gain of non-target position in Fig. 3 (b) is also larger than that in Fig. 3 (d).

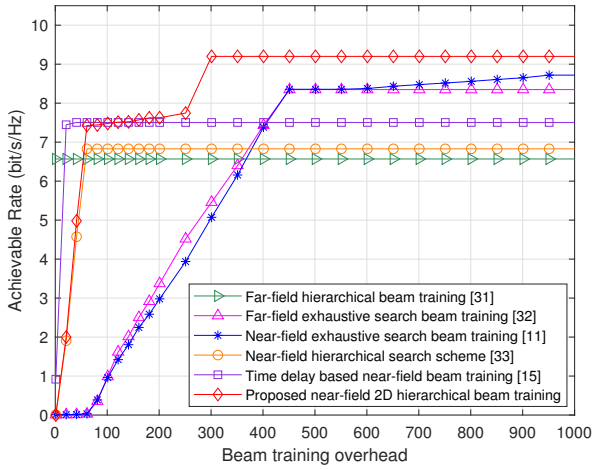
Table. I presents the comparison of beam training overhead for different methods. We compare the proposed near-field 2D hierarchical beam training algorithm with the existing far-field hierarchical beam training scheme [31], far-field exhaustive

TABLE I
COMPARISONS OF BEAM TRAINING OVERHEAD

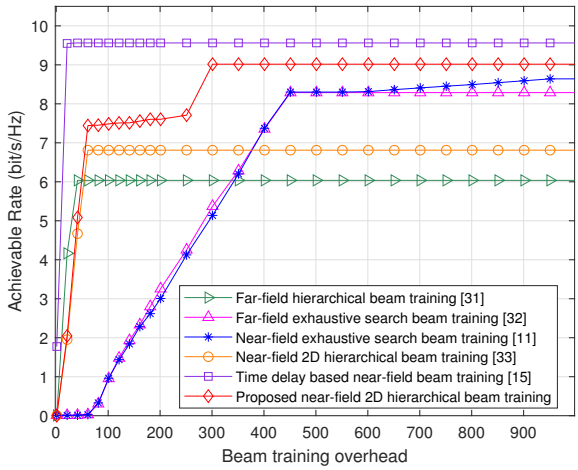
Method	Overhead	Value
Far-field hierarchical scheme [31]	$\sum_l^L U^{(l)}$	40
Far-field exhaustive search scheme [32]	U	512
Near-field exhaustive search scheme [11]	US	8192
Near-field hierarchical search scheme [33]	$\left(\sum_l^L U^{(l)}\right) S$	640
Time-delay based near-field scheme [15]	S	16
Proposed near-field 2D hierarchical scheme	$\sum_l^L U^{(l)} S^{(l)}$	268

search beam training scheme [32], the near-field exhaustive search beam training scheme [11], the existing near-field hierarchical search beam training scheme [33], and time-delay based near-field beam training scheme [15]. We set the number of angle and distance grids as $U = 512$ and $S = 16$, respectively. The overhead of the far-field exhaustive search is set as the same as the number of sampled angle grids, i.e., 512. The overhead of the near-field exhaustive search beam training scheme is set as $512 \times 16 = 8192$. The overhead of time-delay based near-field beam training relates to the number of sampled distance grids, which is set as 16. For the far-field hierarchical beam training scheme, $U^{(l)}$ is the number of sampled angles in the l-th layer, where $U^{(1)} = 4, U^{(2)} = 4, U^{(2)} = 32$. Thus, the overhead of far-field hierarchical beam training is $\sum_l^L U^{(l)} = 4 + 4 + 32 = 40$. Similarly, for near-field hierarchical beam training scheme, the overhead is $40 \times 16 = 640$. For the proposed near-field 2D hierarchical beam training algorithm, we use a three-layer hierarchical codebook. The size of the layer 1 codebook can be calculated as $64 \times 4 = 256$, where the numbers of sampled angle and distance grids are set as 64 and 4. For the layer 2 and layer 3 codebooks, we only need to search 8 and 4 codewords. Thus the overhead of the proposed near-field 2D hierarchical beam training algorithm is 268, which is almost half of 512 and only 3.3 % of 8192.

Fig. 4 presents the performance of achievable rate comparisons against the beam training overhead under different bandwidths. The reason why do the simulation in different bandwidth conditions is that one of the comparisons is time-delay based near-field beam training [15], whose performance heavily lies on the bandwidth. The training overhead increases from 0 to 1000. In the beam training process, we utilize the optimal beamforming vector with the largest achievable rate searched in the current time slots to serve the user. From Fig. 4 (a), where the bandwidth is 100 MHz, we can observe that the proposed near-field 2D hierarchical beam training can achieve the best performance of all schemes with relatively lower overhead. For example, the proposed scheme outperforms the far-field angle-domain codebook with only half of the beam training overhead. The reason is that the existing far-field codebook can only capture the angle information of the channel path. Furthermore, the proposed scheme outperforms the existing near-field hierarchical search scheme. The reason is that the existing near-field beam training scheme does not conduct the optimization of low-resolution codewords to raise the orthogonality of the low-resolution codebook. Moreover, the time-delay based scheme has worse



(a)

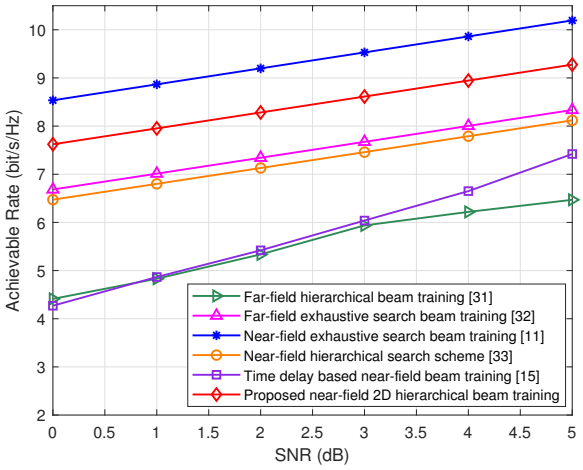


(b)

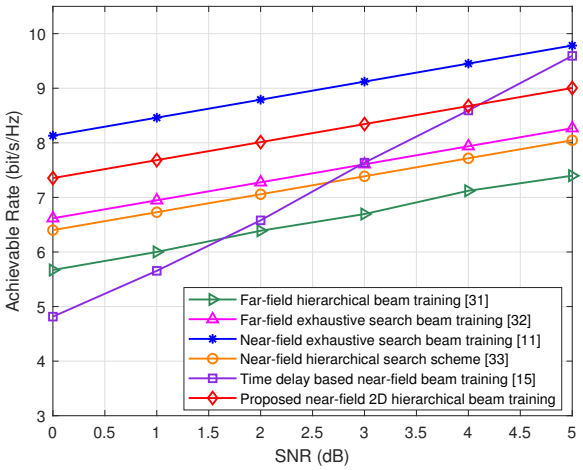
Fig. 4. Achievable sum-rate performance comparison with respect to the beam training overhead under different bandwidths. (a) 100 MHz; (b) 500 MHz.

performance than the proposed scheme in this narrow-band condition. The principal reason is that the ability of time-delay circuits to control the beam split will decrease by reducing the bandwidth. The number of sub-carriers can not support the multiple locations search in one time slot. Meanwhile, Fig. 4 (b) illustrates the wideband situation, where the bandwidth is 500 MHz. It can be observed that the time-delay based beam training scheme has better performance than the proposed scheme since the time-delay based scheme can utilize the beam split effect to search multiple locations in one time slot. However, the proposed scheme has much lower hardware cost and is bandwidth-independent. Thus, we believe that the proposed scheme provides a tradeoff between the performance and overhead in near-field XL-MIMO beam training in a more general and cost-saving way.

Fig. 5 presents the performance of achievable rate comparisons against the SNR under different bandwidths, where SNR is from 0 dB to 5 dB. The simulation parameters are the same as those in Fig. 4. From Fig. 5 (a), i.e., narrow band condition, it is obvious that the proposed beam training scheme

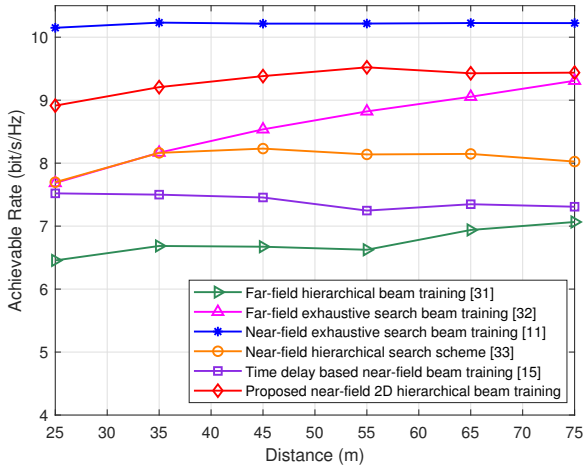


(a)

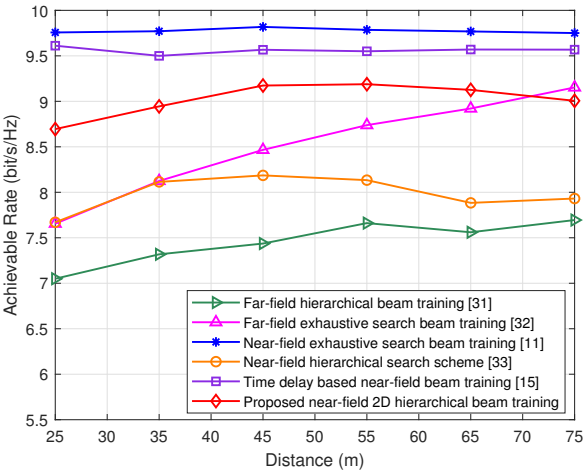


(b)

Fig. 5. Achievable sum-rate performance comparison with respect to the SNR under different bandwidths. (a) 100 MHz; (b) 500 MHz.



(a)



(b)

Fig. 6. Achievable sum-rate performance comparison with respect to the distance overhead under different bandwidths. (a) 100 MHz; (b) 500 MHz.

outperforms all existing far-field and near-field schemes. In specific, around 36.6% improvement in achievable rate is accomplished by the proposed method compared to the time-delay based near-field beam training in SNR = 2 dB. In addition, we can observe that the proposed method can also achieve better performance as long as SNR is smaller than 4 dB in the wideband situation. It is because that near-field beam training scheme is vulnerable to noise is that the time-delay based near-field beam training scheme has to utilize beams with different frequencies to search different locations. the time-delay based near-field beam training scheme can not accumulate the power from all frequencies to combat noise as the near-field exhaustive beam training approach.

Fig. 6 presents the performance of achievable rate comparisons against the distance under different bandwidths, where the distance is from 25 m to 75 m at SNR = 5 dB. From Fig. 6 (a), about 18.5% performance improvement compared to the time-delay based near-field beam training at distance = 55 m. Additionally, we can observe that the proposed method can also reach a 95.8% achievable rate of the time-delay based near-field beam training at distance = 55 m in the wideband situation.

VII. Conclusions

In this paper, we proposed a low-overhead near-field 2D hierarchical beam training by designing the near-field multi-resolution codebooks. Specifically, we first formulate the problem of designing near-field codeword and generating multi-resolution codebooks. It is worth pointing out that the proposed Gerchberg–Saxton (GS) based near-field codeword design algorithm can be utilized in designing codewords to realize arbitrary beam patterns. Then, a low-overhead near-field 2D hierarchical beam training scheme is proposed to realize the tradeoff between the training overhead and performance. Significantly, the proposed scheme can achieve sub-optimal performance without restriction to the hardware cost and wideband condition. In the future, the uniform planar array (UPA) based 3D near-field XL-MIMO beam training is also an attractive research interest.

APPENDIX A
PROOF OF LEMMA 1

The $\|\hat{\mathbf{v}}_{(s)}(u,w)-\hat{\mathbf{v}}'_{(s)}(u,w)\|_2^2$ in (16) can be further presented as

$$\begin{aligned}&\|\hat{\mathbf{v}}_{(s)}(u,w)-\hat{\mathbf{v}}'_{(s)}(u,w)\|_2^2\\=&\|\hat{\mathbf{v}}_{(s)}(u,w)\|_2^2+\|\hat{\mathbf{v}}'_{(s)}(u,w)\|_2^2\\&-2\mathrm{Re}\left(\left\langle\hat{\mathbf{v}}_{(s)}(u,w),\hat{\mathbf{v}}_{(s+1)}(u,w)\right\rangle\right)\\=&\|\hat{\mathbf{v}}_{(s)}(u,w)\|_2^2+\|\hat{\mathbf{v}}'_{(s)}(u,w)\|_2^2\\&-2\mathrm{Re}\left(\|\hat{\mathbf{v}}'_{(s)}(u,w)\|_2\|\hat{\mathbf{v}}_{(s)}(u,w)\|_2\cos\phi\right)\end{aligned}\tag{30}$$

where ϕ is the angle between the $\hat{\mathbf{v}}_{(s)}(u,w)$ and $\hat{\mathbf{v}}'_{(s)}(u,w)$. Meanwhile, $\|\hat{\mathbf{v}}_{(s+1)}(u,w)-\hat{\mathbf{v}}'_{(s)}(u,w)\|_2^2$ can be presented as

$$\begin{aligned}&\|\hat{\mathbf{v}}_{(s+1)}(u,w)-\hat{\mathbf{v}}'_{(s)}(u,w)\|_2^2\\=&\|\hat{\mathbf{v}}_{(s+1)}(u,w)\|_2^2+\|\hat{\mathbf{v}}'_{(s)}(u,w)\|_2^2\\&-2\|\hat{\mathbf{v}}_{(s+1)}(u,w)\|_2\|\hat{\mathbf{v}}'_{(s)}(u,w)\|_2\cos\tau.\end{aligned}\tag{31}$$

where τ is the angle between the $\hat{\mathbf{v}}_{(s+1)}(u,w)$ and $\hat{\mathbf{v}}'_{(s)}(u,w)$. As shown in Step 8 of the **Algorithm 1**, where (12) presents the normalization of the $\hat{\mathbf{v}}'_{(s)}(u,w)$, thus, $\|\hat{\mathbf{v}}_{(s)}(u,w)\|_2 = \|\hat{\mathbf{v}}_{(s+1)}(u,w)\|_2 = 1$, and the phase information of $\hat{\mathbf{v}}_{(s+1)}(u,w)$ and $\hat{\mathbf{v}}'_{(s)}(u,w)$ are the same, i.e., the angle between the $\hat{\mathbf{v}}_{(s+1)}(u,w)$ and $\hat{\mathbf{v}}'_{(s)}(u,w)$ τ is 0. In this case, (30)-(31) is written as

$$\begin{aligned}&\|\hat{\mathbf{v}}_{(s)}(u,w)-\hat{\mathbf{v}}'_{(s)}(u,w)\|_2^2-\|\hat{\mathbf{v}}_{(s+1)}(u,w)-\hat{\mathbf{v}}'_{(s)}(u,w)\|_2^2\\=&\|\hat{\mathbf{v}}_{(s)}(u,w)\|_2^2-\|\hat{\mathbf{v}}_{(s+1)}(u,w)\|_2^2\\&-2\mathrm{Re}\left(\|\hat{\mathbf{v}}'_{(s)}(u,w)\|_2\|\hat{\mathbf{v}}_{(s)}(u,w)\|_2\left(\cos\phi-\cos\tau\right)\right)\\=&2\mathrm{Re}\left(\|\hat{\mathbf{v}}'_{(s)}(u,w)\|_2\left(1-\cos\phi\right)\right)\end{aligned}\tag{32}$$

Since $\|\hat{\mathbf{v}}'_{(s)}(u,w)\|_2\left(1-\cos\phi\right)$ is always greater than zero, we can obtain that

$$\|\hat{\mathbf{v}}_{(s)}(u,w)-\hat{\mathbf{v}}'_{(s)}(u,w)\|_2^2>\|\hat{\mathbf{v}}_{(s+1)}(u,w)-\hat{\mathbf{v}}'_{(s)}(u,w)\|_2^2.\tag{33}$$

REFERENCES

[1] M. Giordani, M. Polese, M. Mezzavilla, S. Rangan, and M. Zorzi, "Toward 6G networks: Use cases and technologies," *IEEE Commun.Mag.*, vol. 58, no. 3, pp. 55–61, Mar. 2020.

[2] P. P. Ray, N. Kumar, and M. Guizani, "A vision on 6G-enabled NIB: Requirements, technologies, deployments, and prospects," *IEEE Wireless Commun.*, vol. 28, no. 4, pp. 120–127, May 2021.

[3] E. D. Carvalho, A. Ali, A. Amiri, M. Angjelichinoski, and R. W. Heath, "Non-stationarities in extra-large-scale massive MIMO," *IEEE Wireless Commun.*, vol. 27, no. 4, pp. 74–80, Aug. 2020.

[4] 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, Jan. 2023.

[5] M. Ke, Z. Gao, Y. Wu, X. Gao, and R. Schober, "Compressive sensing-based adaptive active user detection and channel estimation: Massive access meets massive MIMO," *IEEE Trans. Signal Process.*, vol. 68, pp. 764–779, Jan. 2020.

[6] S. H. Lim, S. Kim, B. Shim, and J. W. Choi, "Efficient beam training and sparse channel estimation for millimeter wave communications under mobility," *IEEE Transa. Commun.*, vol. 68, no. 10, pp. 6583–6596, Jul. 2020.

[7] A. Alkhateeb, O. El Ayach, G. Leus, and R. W. Heath, "Channel estimation and hybrid precoding for millimeter wave cellular systems," *IEEE J. Sel. Top. Signal Process.*, vol. 8, no. 5, pp. 831–846, Oct. 2014.

[8] S. Noh, M. D. Zoltowski, and D. J. Love, "Multi-resolution codebook based beamforming sequence design in millimeter-wave systems," in *Proc. IEEE Global Communications Conference (GLOBECOM'15)*, 2015, pp. 1–6.

[9] 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, Dec. 2020.

[10] A. Pizzo, L. Sanguinetti, and T. L. Marzetta, "Fourier plane-wave series expansion for holographic mimo communications," *IEEE Trans. Wireless Commun.*, vol. 21, no. 9, pp. 237–246, Sep. 2022.

[11] 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, Apr. 2022.

[12] 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, Jun. 2022.

[13] 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.

[14] W. Liu, H. Ren, C. Pan, and J. Wang, "Deep learning based beam training for extremely large-scale massive MIMO in near-field domain," *IEEE Commun. Lett.*, vol. 27, no. 1, pp. 170–174, Jan. 2023.

[15] M. Cui, L. Dai, Z. Wang, S. Zhou, and N. Ge, "Near-field rainbow: Wideband beam training for XL-MIMO," *IEEE Trans. Wireless Commun.*, 2023.

[16] V. Boljanovic, H. Yan, C.-C. Lin, S. Mohapatra, D. Heo, S. Gupta, and D. Cabric, "Fast beam training with true-time-delay arrays in wideband millimeter-wave systems," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 68, no. 4, pp. 1727–1739, Apr. 2021.

[17] 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, pp. 170–180, Dec. 2020.

[18] X. Shi, J. Wang, Z. Sun, and J. Song, "Spatial-chirp codebook-based hierarchical beam training for extremely large-scale massive MIMO," *IEEE Trans. Wireless Commun.*, 2023.

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

[20] C. Huang, L. Liu, C. Yuen, and S. Sun, "Iterative channel estimation using LSE and sparse message passing for mmwave MIMO systems," *IEEE Trans. Signal Process.*, vol. 67, no. 1, pp. 245–259, Nov. 2019.

[21] Z. Xiao, P. Xia, and X.-G. Xia, "Codebook design for millimeter-wave channel estimation with hybrid precoding structure," *IEEE Trans. Wireless Commun.*, vol. 16, no. 1, pp. 141–153, Oct. 2017.

[22] Y. Shechtman, Y. C. Eldar, O. Cohen, H. N. Chapman, J. Miao, and M. Segev, "Phase retrieval with application to optical imaging: A contemporary overview," *IEEE Signal Process. Mag.*, vol. 32, no. 3, pp. 87–109, Apr. 2015.

[23] M. S. Heimbeck and H. O. Everitt, "Terahertz digital holographic imaging," *Advances in Opt. Photonics*, vol. 12, no. 1, pp. 1–59, Mar. 2020.

[24] R. W. Gerchberg and W. O. Saxton, "A practical algorithm for the determination of plane from image and diffraction pictures," *Optik*, vol. 35, no. 2, pp. 237–246, Sep. 1972.

[25] O. Bucci, G. Franceschetti, G. Mazzarella, and G. Panariello, "Intersection approach to array pattern synthesis," *IEEE Photonics Journal*, vol. 137, no. 6, pp. 349–357, Dec. 1990.

[26] W. Chen, "3D gerchberg-saxton optical correlation," *IEEE Photonics Journal*, vol. 10, no. 2, pp. 1–9, Apr. 2018.

[27] J. Miao, P. Charalambous, J. Kirz, and D. Sayre, "Extending the methodology of X-ray crystallography to allow imaging of micrometre-sized non-crystalline specimen," *Nature*, vol. 400, no. 6742, pp. 342–344, May 1999.

[28] R. N. Bracewell, *The Fourier Transform and Its Applications*. New York: McGraw-Hill, 1986.

[29] R. A. Horn and C. R. Johnson, *Matrix Analysis*. New York: Cambridge University Press, 2012.

[30] G. Jiang and C. Qi, "Near-field beam training based on deep learning for extremely large-scale MIMO," *IEEE Commun. Lett.*, vol. 27, no. 8, pp. 2063–2067, Jun. 2023.

[31] S. Noh, M. D. Zoltowski, and D. J. Love, "Multi-resolution codebook and adaptive beamforming sequence design for millimeter wave beam alignment," *IEEE Trans. Wireless Commun.*, vol. 16, no. 9, pp. 5689–5701, Sep. 2017.

[32] J. Lee, G. Gil, and Y. H. Lee, "Channel estimation via orthogonal matching pursuit for hybrid MIMO systems in millimeter wave communications," *IEEE Trans. Wireless Commun.*, vol. 64, no. 6, pp. 2370–2386, Jun. 2016.

[33] J. Chen, F. Gao, M. Jian, and W. Yuan, "Hierarchical codebook design for near-field mmwave MIMO communications systems," *IEEE Wireless Commun. Lett.*, 2023.