Beam Refinement for THz Extremely Large-Scale MIMO systems based on Gaussian Approximation

Kangjian Chen *†, Chenhao Qi *† and Octavia A. Dobre ‡
*School of Information Science and Engineering, Southeast University, Nanjing, China
†National Mobile Communications Research Laboratory, Southeast University, Nanjing, China
‡Faculty of Engineering and Applied Science, Memorial University, Canada
Email: {kjchen, qch}@seu.edu.cn, odobre@mun.ca

GLOBECOM 2023 - 2023 IEEE Global Communications Conference | 979-8-3503-1090-0/23/$31.00 ©2023 IEEE | DOI: 10.1109/GLOBECOM54140.2023.10437496

***Abstract*—Beam refinement is a key technology to overcome the problem of limited resolution in beam training. However, most existing works on beam refinement are not suitable for the emerging extremely large-scale multiple-input-multiple-output (XL-MIMO) due to the differences in the channel characteristics. To fill in the gap, in this paper, beam refinement for XL-MIMO systems is investigated. Inspired by the similarities between the Taylor series of the Gaussian function and that of the beam gain, we propose to approximate the beam gain by the Gaussian function. Then, a low-complexity beam refinement based on the Gaussian approximation (BRGA) scheme, which quantizes the narrowed intervals after beam training into several samples and performs additional channel tests on the quantized grids, is proposed to improve the estimation accuracy of the beam training. Based on the measurements in the beam refinement stage, the BRGA-based least square (BRGA-LS) estimator is developed for high-resolution channel parameter estimation. To avoid the noise amplification effects of the BRGA-LS, the BRGA-based weighted least square (BRGA-WLS) estimator is further developed. Simulation results verify the effectiveness of the proposed scheme and show that the proposed BRGA scheme can greatly improve the accuracy of beam training with only a few additional channel tests.**

***Index Terms*—Beam training, beam refinement, extremely large-scale MIMO (XL-MIMO), Gaussian approximation.**

I. Introduction

Terahertz (THz) that reserves wide spectrum resources is a promising technology for achieving high data rates. Thanks to its small wavelength, the space-limited base station (BS) can accommodate a large number of antennas to enhance the spectral efficiency by implementing the massive multiple-input-multiple-output (MIMO). The perfect match between the abundant spectrum resources provided by THz and the high spectral efficiency enabled by massive MIMO has led to the boom of THz massive MIMO [1].

Recently, extremely large-scale MIMO (XL-MIMO) with far more antennas than existing massive MIMO has been developed to further improve the spectral efficiency via ultrahigh-gain beamforming [2]. Attracted by the tremendous potentials, researchers have moved on to develop XL-MIMO based on the well-explored massive MIMO. However, extending the works in massive MIMO to XL-MIMO is not straightforward. To be specific, due to the much larger array aperture, the channel of the latter differs greatly from that of the former in terms of propagation characteristics. Generally, the radiation fields of the electromagnetic waves can be divided into the near field and the far field according to the distance between the BS and the radiation source [3], [4]. In the existing massive MIMO, the coverage of the BS is dominated by the far field, where the electromagnetic wave is approximated as the planar wave and hence the phase differences of the antennas can be modeled as the linear function of the antenna indices. In the emerging XL-MIMO, the near field takes a significant proportion of the BS coverage, where the spherical wave is adopted to characterize the electromagnetic wave propagation and the phase differences of the antennas are the nonlinear function of the antenna indices.

One pivotal research topic of wireless communications is the channel state information (CSI) acquisition [5]. In the context of XL-MIMO, CSI acquisition should take both the near field and the far field into account. Many interesting researches have been conducted for the CSI acquisition of XL-MIMO. In [6], a polar-domain simultaneous orthogonal matching pursuit (P-SOMP) algorithm is proposed to estimate the near-field channels by exploiting the polar domain sparsity. In [7], a chirp-based hierarchical beam training scheme is developed for the XL-MIMO. However, the accuracy of the dictionary-based channel estimation method in [6] and the beam training methods in [7] is limited by the quantization of space. To overcome the limited resolution resulting from the quantization, low-complexity beam refinement is developed. In [8], an auxiliary beam pair (ABP) is proposed for beam refinement, which obtains the high-accuracy estimates of channel angles by comparing the beam gain of the adjacent codewords. In [9], a virtual-subarray-based beam refinement scheme is proposed, where the antenna array is virtually divided into two subarrays, and phase differences of several measurements are exploited for channel angle estimation. In [10], a far-field beam refinement (FFBR) method for analog antenna arrays is developed by approximating the array gain as a Gaussian function. However, the beam refinement methods in [8]–[10] only adapt to the far-field channels. To our best knowledge, there have been no published works on beam refinement for the XL-MIMO considering both the near field and the far field.

In this paper, beam refinement for XL-MIMO systems with hybrid precoding structure is investigated. Inspired by the

similarities between the Taylor series of the Gaussian function and that of the beam gain, we propose to approximate the array gain of XL-MIMO with a two-dimensional Gaussian function. Then, a low-complexity beam refinement based on the Gaussian approximation (BRGA) scheme is proposed to improve the estimation accuracy of the beam training, which quantizes the narrowed intervals after beam training into several samples and performs additional channel tests on the quantized grids. Based on the measurements in the beam refinement stage, the BRGA-based least square (BRGA-LS) estimator is developed for high-resolution channel parameter estimation. To avoid the noise amplification effects of the BRGA-LS, the BRGA-based weighted least square (BRGA-WLS) estimator is further developed.

The notations are defined as follows. Symbols for matrices (upper case) and vectors (lower case) are in boldface. $(\cdot)^{\mathrm{T}}$ and $(\cdot)^{\mathrm{H}}$ denote the transpose and conjugate transpose (Hermitian), respectively. $[\boldsymbol{a}]_n$, $[\boldsymbol{A}]_{:,n}$ and $[\boldsymbol{A}]_{m,n}$ denote the nth entry of vector $\boldsymbol{a}$, the nth column of matrix $\boldsymbol{A}$ and the entry on the mth row and the nth column of matrix $\boldsymbol{A}$, respectively. In addition, j, $|\cdot|$, $\mathbb{C}$, and $\mathcal{CN}$ denote the square root of -1, the absolute value of a scalar, the set of the complex number and the complex Gaussian distribution.

II. System Model

As shown in Fig. 1, we consider the uplink beam training between the BS and the user, where a half-wavelength-interval array with N_{BS} antennas is equipped at the BS while a single-antenna transceiver is adopted for the user. To ease the notation, we assume N_{BS} is an odd number and $N_{\mathrm{BS}} = 2N+1$. With the hybrid combining structure, the N_{BS} antennas are connected to N_{RF} RF chains through the phase shifter network. In this work, the beam training is performed based on the analog combining, and hence we focus on one of the multiple RF chains for simplicity. Then the pth received signal, for $p = 1, 2, \cdots P$, can be expressed as

$$y_p = \boldsymbol{w}_p^{\mathrm{H}} \boldsymbol{h} x_p + \boldsymbol{w}_p^{\mathrm{H}} \boldsymbol{\eta} \tag{1}$$

where $\boldsymbol{w}_p \in \mathbb{C}^{N_{\mathrm{BS}}}$, $\boldsymbol{h} \in \mathbb{C}^{N_{\mathrm{BS}}}$, and x_p denote the analog combiner at the BS, the channel between the BS and the user, and the transmit signal, respectively. In addition, $\boldsymbol{\eta} \sim \mathcal{CN}(\boldsymbol{0}, \varrho_\eta^2 \boldsymbol{I}_{N_{\mathrm{BS}}})$ denotes the additive while Gaussian noise.

To depict the channel, we first establish a Cartesian coordinate, which sets the center, the normal direction, and the tangent direction of the antenna array as the origin, the x-axis, and the y-axis, respectively. Then the coordinate of the nth antenna can be expressed as $(0, n\lambda/2)$ for $n \in \boldsymbol{\mathcal{I}}$, where $\boldsymbol{\mathcal{I}} \triangleq \{-N, \cdots, 0, \cdots, N\}$ and λ denotes the wavelength. From Fig. 1, the coordinate of the radiation source (the user or scatterers) at the lth path is $(r_l \cos\theta_l, r_l \sin\theta_l)$, for $l = 1, 2, \cdots, L$, where L denotes the number of paths, r_l denotes the distance between the origin and the lth radiation source, and $\theta_l \in [-\pi/2, \pi/2]$ denotes the angle of the lth radiation source relative to the x-axis. The distance between the lth radiation source and the nth antenna is calculated as

$$r_l^{(n)} = \sqrt{r_l^2 + n^2\lambda^2/4 - n r_l \Theta_l \lambda} \tag{2}$$

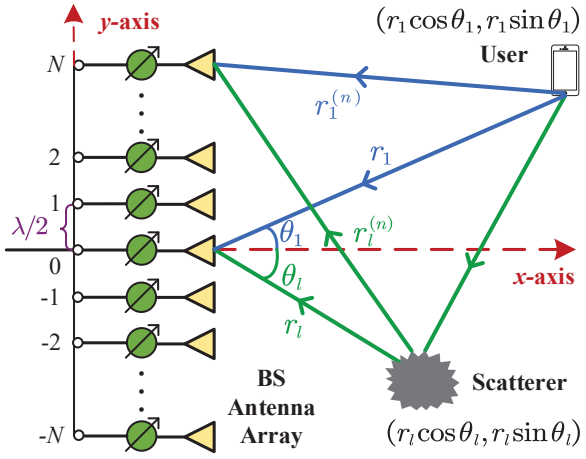

Fig. 1. Illustration of the system model.

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

$$\boldsymbol{h} = \sum_{l=1}^{L} g_l \boldsymbol{\alpha}(\Theta_l, r_l) \tag{3}$$

where g_l denotes the channel gain of the lth path. The channel steering vector $\boldsymbol{\alpha}(\cdot)$ is defined as

$$\boldsymbol{\alpha}(\Theta_l, r_l) = \frac{1}{\sqrt{N_{\mathrm{BS}}}} \left[e^{-j\frac{2\pi}{\lambda}(r_l^{(-N)} - r_l)}, \ldots, e^{-j\frac{2\pi}{\lambda}(r_l^{(N)} - r_l)} \right]^{\mathrm{T}}. \tag{4}$$

III. Beam Refinement Based on Gaussian Approximation

In this section, we focus on the beam refinement for XL-MIMO. First of all, we introduce the codebook design and the beam training for XL-MIMO. Then, the similarities between the Taylor series of the Gaussian function and that of the beam gain are presented. In addition, a BRGA scheme is proposed, which quantizes the narrowed intervals after beam training into several samples and performs additional channel tests on the quantized grids. Based on the measurements in the beam refinement stage, two estimators including BRGA-LS and BRGA-WLS are developed for high-resolution channel parameter estimation.

A. Codebook Design and Beam Training

We omit the subscript "l" of $r_l^{(n)}$ and simplify $r^{(n)}$ as

$$r^{(n)} \approx r - n\Theta\lambda/2 + \frac{n^2\lambda^2(1-\Theta^2)}{8r} \tag{5}$$

according to $\sqrt{1+\epsilon} \approx 1 + \frac{1}{2}\epsilon - \frac{1}{8}\epsilon^2$, which is verified to be accurate if $r^{(n)} \geq 0.5\sqrt{N^3\lambda^2}$ [4], [6]. In fact, the boundary $0.5\sqrt{N^3\lambda^2}$ is quite small compared to the coverage of the BS. For example, if $N = 128$ and $\lambda = 0.003$ m, $0.5\sqrt{N^3\lambda^2} \approx 2.2$ m, which is much smaller than the typical coverage of the BS. Therefore, in this work, we focus on the radiation field

with $r^{(n)} \geq 0.5\sqrt{N^3\lambda^2}$. Substituting (5) into (4), we have

$$[\boldsymbol{\alpha}(\Theta, r)]_n \approx \frac{1}{\sqrt{N_{\mathrm{BS}}}} e^{j\pi\left(\Theta n - \frac{\lambda(1-\Theta^2)}{4r} n^2\right)} \tag{6}$$

for $n \in \boldsymbol{\mathcal{I}}$. Define

$$k \triangleq \frac{\lambda(1-\Theta^2)}{4r}. \tag{7}$$

Then (4) can be approximated as

$$\begin{aligned}\boldsymbol{\alpha}(\Theta, r) &\approx \frac{1}{\sqrt{N_{\mathrm{BS}}}} \left[e^{j\pi(-\Theta N - kN^2)}, \ldots, e^{j\pi(\Theta N - kN^2)}\right]^{\mathrm{T}} \\ &\triangleq \boldsymbol{\gamma}(\Theta, k).\end{aligned} \tag{8}$$

From (8), the channel steering vector is related to both Θ and k. To establish a codebook covering the whole space, we quantize Θ by Q samples and quantize k by S samples. Since $\Theta \in [-1, 1]$, the qth sample of Θ is expressed as

$$\varphi_q = (2q - 1 - Q)/Q \tag{9}$$

for $q = 1, 2, \cdots, Q$. Note that $0 \leq k \leq \frac{1}{2\sqrt{N^3}}$ because $r \geq 0.5\sqrt{N^3\lambda^2}$. Therefore, the sth sample of k is expressed as

$$t_s = \frac{(s-1)}{2\sqrt{N^3}(S-1)} \tag{10}$$

for $s = 1, 2, \cdots S$. Then the codebook can be established as

$$\boldsymbol{C} \triangleq \{\boldsymbol{C}_1, \boldsymbol{C}_2, \cdots \boldsymbol{C}_Q\} \in \mathbb{C}^{N \times QS} \tag{11}$$

where $\boldsymbol{C}_q \in \mathbb{C}^{N \times S}$ and $[\boldsymbol{C}_q]_{:,s} = \boldsymbol{\gamma}(\varphi_q, t_s)$ for $q = 1, 2, \cdots, Q$ and $s = 1, 2, \cdots S$.

To find the LoS path, the codewords in $\boldsymbol{C}$ need to be tested one by one [11]. Denote $\boldsymbol{Y} \in \mathbb{C}^{Q \times S}$ as the matrix to keep the received signals for the beam training. When testing the sth codeword in $\boldsymbol{C}_q$, the received signal at the BS can be expressed as

$$[\boldsymbol{Y}]_{q,s} = [\boldsymbol{C}_q]_{:,s}^{\mathrm{H}} \boldsymbol{h} + [\boldsymbol{C}_q]_{:,s}^{\mathrm{H}} \boldsymbol{\eta} \tag{12}$$

where we set the transmit symbol as "1". Then the indices of the codeword in $\boldsymbol{C}$ best fitting for $\boldsymbol{h}$ are determined by [12]

$$\begin{aligned}&(\widehat{q}, \widehat{s}) = \arg \max_{q,s} \ |[\boldsymbol{Y}]_{q,s}| \\ &\text{s.t. } q = 1, \cdots Q, \ s = 1, \cdots, S, \ p = (q-1)S + s.\end{aligned} \tag{13}$$

B. Gaussian Approximation

For an arbitrary channel steering vector, $\boldsymbol{v} = \boldsymbol{\gamma}(\Theta, k)$, we define its hybrid-field beam gain as

$$\begin{aligned}G(\boldsymbol{v}, \Omega, b) &\triangleq N_{\mathrm{BS}} \boldsymbol{\gamma}(\Omega, b)^{\mathrm{H}} \boldsymbol{v} \\ &= \sum_{n=-N}^{N} e^{j\pi((\Theta-\Omega)n - (k-b)n^2)} \\ &= \sum_{n=-N}^{N} e^{j(\widetilde{\Omega} n - \widetilde{b} n^2)}\end{aligned} \tag{14}$$

where $\widetilde{\Omega} \triangleq (\Theta - \Omega)\pi$ and $\widetilde{b} \triangleq (k - b)\pi$. Note that different from the beam gain of the channel steering vector for massive MIMO that only relates to the spatial angle, the hybrid-field beam gain of the channel steering vector for XL-MIMO is related to both the spatial angle and the distance. Furthermore, we have

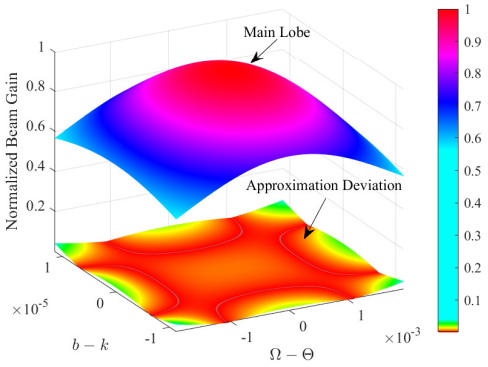

Fig. 2. Illustration of the Gaussian Approximation.

$$\begin{aligned}|G(\boldsymbol{v}, \Omega, b)|^2 &= \left(\sum_{n=-N}^{N} e^{j(\widetilde{\Omega} n - \widetilde{b} n^2)}\right)\left(\sum_{n=-N}^{N} e^{j\pi(\widetilde{\Omega} n - \widetilde{b} n^2)}\right)^* \\ &= \sum_{n=-N}^{N}\sum_{m=-N}^{N} e^{j\left(\widetilde{\Omega}(n-m) - \widetilde{b}(n^2-m^2)\right)} \\ &= \sum_{n=-N}^{N}\sum_{m=-N}^{N} \cos\left(\widetilde{\Omega}(n-m) + \widetilde{b}(m^2-n^2)\right).\end{aligned} \tag{15}$$

Unfortunately, it is hard to obtain the closed-form solution of (15) due to the quadratic phase term $\widetilde{b}(m^2 - n^2)$. To obtain a deeper scope of $|G(\boldsymbol{v}, \Omega, b)|^2$, we resort to the widely-used Taylor series.

At the top of the next page, we provide the Taylor series of $|G(\boldsymbol{v}, \Omega, b)|^2$ and $e^{-\widetilde{\Omega}^2/\sigma_1^2 - \widetilde{b}^2/\sigma_2^2}$ about the point $(0, 0)$, where C_p, $\widetilde{C}_p$, $\overline{C}_{p,q}$, D_p, $\widetilde{D}_p$, $\overline{D}_{p,q}$ are all constant coefficients. In addition, we omit lower-order terms in (a) of (16) for simplicity. From (16) and (17), the Taylor series of $|G(\boldsymbol{v}, \Omega, b)|^2$ and $e^{-\widetilde{\Omega}^2/\sigma_1^2 - \widetilde{b}^2/\sigma_2^2}$ share similarities in several aspects including the orders and coefficients of the series. Inspired by the similarities, we propose to use the two-dimensional Gaussian function

$$f(\Omega, b) \triangleq a e^{-\frac{(\Omega-\Theta)^2}{2\sigma_1^2} - \frac{(b-k)^2}{2\sigma_2^2}}. \tag{18}$$

to approximate the main lobe of $|G(\boldsymbol{v}, \Omega, b)|$, which is the Gaussian approximation. From (9) and (10), the quantization intervals of Θ and k are $\Delta\Theta = 2/Q$ and $\Delta k = \frac{1}{2\sqrt{N^3}(S-1)}$, respectively. Therefore, the main lobe of $|G(\boldsymbol{v}, \Omega, b)|$ is restricted to the region

$$\left\{(\Omega, b) \middle| \Omega \in \left[\Theta - \frac{\Delta\Theta}{2}, \Theta + \frac{\Delta\Theta}{2}\right], \ b \in \left[k - \frac{\Delta k}{2}, k + \frac{\Delta k}{2}\right]\right\}. \tag{19}$$

Then, the Gaussian approximation is formulated as

$$\min_{a,\sigma_1,\sigma_2} \int_{k-\frac{\Delta k}{2}}^{k+\frac{\Delta k}{2}} \int_{\Theta-\frac{\Delta\Theta}{2}}^{\Theta+\frac{\Delta\Theta}{2}} \left| f(\Omega, b) - |G(\boldsymbol{v}, \Omega, b)| \right|^2 \mathrm{d}\Omega \mathrm{d}b \tag{20}$$

$$|G(\boldsymbol{v},\Omega,b)|^2=\sum_{n=-N}^{N}\sum_{m=-N}^{N}\sum_{p=0}^{\infty}\sum_{q=1}^{2p-1}\frac{(-1)^p(n-m)^{2p}\widetilde{\Omega}^{2p}}{(2p)!}+\frac{(-1)^p(m^2-n^2)^{2p}\widetilde{b}^{2q}}{(2p)!}+2\frac{(-1)^p(n-m)^q(m^2-n^2)^{2p-q}\widetilde{\Omega}^q\widetilde{b}^{2p-q}}{(q)!(2p-q)!}$$
$$\overset{(a)}{\approx}\sum_{p=0}^{\infty}\sum_{q=1}^{2p-1}\frac{(-1)^pC_pN^{2p}\widetilde{\Omega}^{2p}}{(2p)!}+\frac{(-1)^p\widetilde{C}_p(N^2)^{2p}\widetilde{b}^{2q}}{(2p)!}+2\frac{(-1)^p\overline{C}_{p,q}N^{4p-q+2}\widetilde{\Omega}^q\widetilde{b}^{2p-q}}{(q)!(2p-q)!} \tag{16}$$

$$e^{-\frac{\widetilde{\Omega}^2}{\sigma_1^2}-\frac{\widetilde{b}^2}{\sigma_2^2}}=\sum_{p=0}^{\infty}\sum_{q=1}^{p-1}\frac{(-1)^p\left(\frac{2}{\sigma_1^2}\right)^p\widetilde{\Omega}^{2p}\prod_{i=1}^{p}(2i-1)}{(2p)!}+\frac{(-1)^p\left(\frac{2}{\sigma_2^2}\right)^p\widetilde{b}^{2q}\prod_{i=1}^{p}(2i-1)}{(2p)!}$$
$$+2\frac{(-1)^p\left(\frac{2}{\sigma_1^2}\right)^q\left(\frac{2}{\sigma_2^2}\right)^{p-q}\left(\prod_{i=1}^{q}(2i-1)\right)\left(\prod_{i=1}^{p-q}(2i-1)\right)\widetilde{\Omega}^{2q}\widetilde{b}^{2p-2q}}{(2q)!(2p-2q)!}$$
$$=\sum_{p=0}^{\infty}\sum_{q=1}^{p-1}\frac{(-1)^pD_p\left(\sigma_1^{-1}\right)^{2p}\widetilde{\Omega}^{2p}}{(2p)!}+\frac{(-1)^p\widetilde{D}_p\left(\sigma_2^{-1}\right)^{2p}\widetilde{b}^{2q}}{(2p)!}+2\frac{(-1)^p\overline{D}_{p,q}\left(\sigma_1^{-1}\right)^{2q}\left(\sigma_2^{-1}\right)^{2p-2q}\widetilde{\Omega}^{2q}\widetilde{b}^{2p-2q}}{(2q)!(2p-2q)!} \tag{17}$$

which is a nonlinear least-square problem and can be solved by the trust-region optimization algorithm [13]. We omit the details and denote the solutions of (20) as $\widehat{a}$, $\widehat{\sigma}_1$ and $\widehat{\sigma}_2$. Then, the optimized Gaussian function can be expressed as

$$\widehat{f}(\Omega,b)=\widehat{a}e^{-\frac{(\Omega-\Theta)^2}{2\widehat{\sigma}_1^2}-\frac{(b-k)^2}{2\widehat{\sigma}_2^2}}. \tag{21}$$

Note that for another channel steering vector, $\overline{\boldsymbol{v}}=\boldsymbol{\gamma}(\overline{\Theta},\overline{k})$, we have

$$G(\overline{\boldsymbol{v}},\Omega,b)=\sum_{n=-N}^{N}e^{j\pi((\overline{\Theta}-\Omega)n-(\overline{k}-b)n^2)}$$
$$=\sum_{n=-N}^{N}e^{j\pi((\Theta-(\Theta-\overline{\Theta}+\Omega))n-(k-(k-\overline{k}+b))n^2)}$$
$$=G(\boldsymbol{v},\Omega+(\Theta-\overline{\Theta}),b+(k-\overline{k})) \tag{22}$$

which indicates that the beam gain of $\overline{\boldsymbol{v}}$ is the translation of that of $\boldsymbol{v}$. Therefore, we only need to solve (20) for $\boldsymbol{v}$, and the Gaussian approximation for other channel steering vectors can be obtained via the translation in (22).

As shown in Fig. 2, we illustrate the main lobe of $|G(\boldsymbol{v},\Omega,b)|$ and the deviation of the Gaussian approximation, where we set $N=256$, $Q=513$, and $S=6$. From the figure, the approximation deviation is quite small compared to beam gain of the main lobe. For example, the maximum approximation deviation is only 4% of the maximum beam gain and the averaged approximation deviation is only 0.5% of the maximum beam gain.

C. Beam Refinement based on Gaussian Approximation

Note that the estimation accuracy of (13) is limited by the quantization intervals of Θ and k. A straightforward way to improve the estimation accuracy is to reduce the quantization intervals. However, the smaller quantization intervals will inevitably lead to a larger codebook size, and higher training overheads are needed to test the codewords in the enlarged codebook. To improve the estimation accuracy of beam training results with low overhead, we then propose a BRGA scheme.

First, we determine the limits of Θ and k based on the beam training results. From (13), the indices of the beam training results are $\widehat{q}$ and $\widehat{s}$, respectively. Therefore, the limits of Θ and k can be expressed as

$$\boldsymbol{\Theta}=\left[\varphi_{\widehat{q}}-\frac{\Delta\Theta}{2},\varphi_{\widehat{q}}+\frac{\Delta\Theta}{2}\right],$$
$$\boldsymbol{\mathcal{K}}=\left[t_{\widehat{s}}-\frac{\Delta k}{2},t_{\widehat{s}}+\frac{\Delta k}{2}\right]. \tag{23}$$

Then, we further narrow down the limits of Θ and k by comparing the powers of the received signals adjacent to $[\boldsymbol{Y}]_{\widehat{q},\widehat{s}}$. The narrowed limits of Θ are expressed as

$$\widetilde{\boldsymbol{\Theta}}=\begin{cases}\left[\varphi_{\widehat{q}},\varphi_{\widehat{q}}+\frac{\Delta\Theta}{2}\right], & [\boldsymbol{Y}]_{\widehat{q}+1,\widehat{s}}\geq[\boldsymbol{Y}]_{\widehat{q}-1,\widehat{s}},\\ \left[\varphi_{\widehat{q}}-\frac{\Delta\Theta}{2},\varphi_{\widehat{q}}\right], & [\boldsymbol{Y}]_{\widehat{q}+1,\widehat{s}}<[\boldsymbol{Y}]_{\widehat{q}-1,\widehat{s}}.\end{cases} \tag{24}$$

Similarly, the narrowed limits of k can be expressed as

$$\widetilde{\boldsymbol{\mathcal{K}}}=\begin{cases}\left[t_{\widehat{s}},t_{\widehat{s}}+\frac{\Delta k}{2}\right], & [\boldsymbol{Y}]_{\widehat{q},\widehat{s}+1}\geq[\boldsymbol{Y}]_{\widehat{q},\widehat{s}-1},\\ \left[t_{\widehat{s}}-\frac{\Delta k}{2},t_{\widehat{s}}\right], & [\boldsymbol{Y}]_{\widehat{q},\widehat{s}+1}<[\boldsymbol{Y}]_{\widehat{q},\widehat{s}-1}.\end{cases} \tag{25}$$

After that, we quantize the narrowed limits of Θ and k by M samples. Denote the left and right boundaries of $\widetilde{\boldsymbol{\Theta}}$ as $\widetilde{\Theta}_{\mathrm{L}}$ and $\widetilde{\Theta}_{\mathrm{R}}$, respectively. Similarly, we have $\widetilde{\boldsymbol{\mathcal{K}}}=[\widetilde{k}_{\mathrm{L}},\widetilde{k}_{\mathrm{R}}]$. Then the mth sample of Θ and k can be expressed as

$$\widetilde{\Theta}_m=\widetilde{\Theta}_{\mathrm{L}}+\frac{(m-1)(\widetilde{\Theta}_{\mathrm{R}}-\widetilde{\Theta}_{\mathrm{L}})}{M-1},$$
$$\widetilde{k}_m=\widetilde{k}_{\mathrm{L}}+\frac{(m-1)(\widetilde{k}_{\mathrm{R}}-\widetilde{k}_{\mathrm{L}})}{M-1}. \tag{26}$$

To perform the beam refinement, the user continues to transmit pilots. Meanwhile, the BS receives the pilots via $\boldsymbol{\gamma}(\widetilde{\Theta}_m,\widetilde{k}_t)$ for $m=1,2,\cdots,M$ and $t=1,2,\cdots,M$. Then the received signals at the BS can be expressed as

$$\widetilde{y}_{m,t}=\boldsymbol{\gamma}(\widetilde{\Theta}_m,\widetilde{k}_t)^{\mathrm{H}}\boldsymbol{h}+\boldsymbol{\gamma}(\widetilde{\Theta}_m,\widetilde{k}_t)^{\mathrm{H}}\boldsymbol{\eta}$$
$$\overset{(a)}{\approx}g\boldsymbol{\gamma}(\widetilde{\Theta}_m,\widetilde{k}_t)^{\mathrm{H}}\boldsymbol{\alpha}(\Theta_1,r_1)+\eta$$
$$\overset{(b)}{\approx}g\boldsymbol{\gamma}(\widetilde{\Theta}_m,\widetilde{k}_t)^{\mathrm{H}}\boldsymbol{\gamma}(\Theta,k)+\eta \tag{27}$$

where we omit the NLoS channels in (a), omit the subscript "1" and approximate $\boldsymbol{\alpha}(\Theta, r)$ with $\boldsymbol{\gamma}(\Theta, k)$ in (b). In (27), $\eta \triangleq \boldsymbol{\gamma}(\widetilde{\Theta}_m, \widetilde{k}_t)^{\mathrm{H}}\boldsymbol{\eta}$ denotes the noise term. In addition, we specify that all the variants of η denote the noise terms in the following text.

Applying the Gaussian approximation to (27), we have

$$\begin{aligned}|\widetilde{y}_{m,t}| &\approx |\widetilde{g}|\widehat{f}(\widetilde{\Theta}_m, \widetilde{k}_t) + \widetilde{\eta} \\ &= |\widetilde{g}|\widehat{a}e^{-\frac{(\widetilde{\Theta}_m-\Theta)^2}{2\widehat{\sigma}_1^2} - \frac{(\widetilde{k}_t-k)^2}{2\widehat{\sigma}_2^2}} + \widetilde{\eta} \end{aligned} \tag{28}$$

where $\widetilde{g} = g/N_{\mathrm{BS}}$. Note that the Gaussian function is the exponential of a quadratic function. Taking the natural logarithm of $|\widetilde{y}_{m,t}|$, we have

$$\ln|\widetilde{y}_{m,t}| = \ln(|\widetilde{g}|\widehat{a}) - \frac{(\widetilde{\Theta}_m-\Theta)^2}{2\widehat{\sigma}_1^2} - \frac{(\widetilde{k}_t-k)^2}{2\widehat{\sigma}_2^2} + \overline{\eta}. \tag{29}$$

In (29), we establish a quadratic function about the channel parameters Θ and k. We then develop two estimators termed as BRGA-LS and BRGA-WLS to obtain the high-resolution estimates of Θ and k based on (29).

1) BRGA-LS: To remove the constant and quadratic terms, we subtract $\ln|\widetilde{y}_{1,1}|$ from $\ln|\widetilde{y}_{m,t}|$ and obtain

$$\ln|\widetilde{y}_{m,t}| - \ln|\widetilde{y}_{1,1}| = \frac{(\widetilde{\Theta}_m-\widetilde{\Theta}_1)}{\widehat{\sigma}_1^2}\Theta + \frac{(\widetilde{k}_t-\widetilde{k}_1)}{\widehat{\sigma}_2^2}k + \overline{\overline{\eta}}. \tag{30}$$

We rewrite (30) in a more compact form as

$$\boldsymbol{A}_{\mathrm{LS}}\boldsymbol{z}_{\mathrm{LS}} + \overline{\overline{\boldsymbol{\eta}}} = \boldsymbol{y}_{\mathrm{LS}} \tag{31}$$

where $\boldsymbol{z}_{\mathrm{LS}} \triangleq [\Theta, k]^{\mathrm{T}}$ and $\overline{\overline{\boldsymbol{\eta}}}$ is the stack of noise terms. The uth row of $\boldsymbol{A}_{\mathrm{LS}}$ and the uth entry of $\boldsymbol{y}_{\mathrm{LS}}$ are expressed as

$$\begin{aligned}[\boldsymbol{A}_{\mathrm{LS}}]_{u,:} &= \left[\frac{(\widetilde{\Theta}_m-\widetilde{\Theta}_1)}{\widehat{\sigma}_1^2}, \frac{(\widetilde{k}_t-\widetilde{k}_1)}{\widehat{\sigma}_2^2}\right] \\ [\boldsymbol{y}_{\mathrm{LS}}]_u &= \ln|\widetilde{y}_{m,t}| - \ln|\widetilde{y}_{1,1}| \end{aligned} \tag{32}$$

for $u = (m-1)M + t$, $m = 1, 2, \cdots, M$, $t = 1, 2, \cdots, M$ and $m + t \neq 2$. The least-square solution of (31) can be expressed as

$$\widehat{\boldsymbol{z}}_{\mathrm{LS}} = (\boldsymbol{A}_{\mathrm{LS}}^{\mathrm{T}}\boldsymbol{A}_{\mathrm{LS}})^{-1}\boldsymbol{A}_{\mathrm{LS}}^{\mathrm{T}}\boldsymbol{y}_{\mathrm{LS}}. \tag{33}$$

Then the estimation of Θ and k can be expressed as

$$\widehat{\Theta}_{\mathrm{LS}} = [\widehat{\boldsymbol{z}}_{\mathrm{LS}}]_1, \ \widehat{k}_{\mathrm{LS}} = [\widehat{\boldsymbol{z}}_{\mathrm{LS}}]_2. \tag{34}$$

From (7), the estimation of r can be expressed as

$$\widehat{r}_{\mathrm{LS}} = \frac{\lambda(1-\widehat{\Theta}_{\mathrm{LS}}^2)}{4\widehat{k}_{\mathrm{LS}}}. \tag{35}$$

2) BRGA-WLS: From (28), we have

$$\begin{aligned}\ln|\widetilde{y}_{m,t}| &\approx \ln(|\widetilde{g}|\widehat{f}(\widetilde{\Theta}_m, \widetilde{k}_t)) + \ln\left(1 + \frac{\widetilde{\eta}}{|\widetilde{g}|\widehat{f}(\widetilde{\Theta}_m, \widetilde{k}_t)}\right) \\ &\overset{(a)}{\approx} \ln(|\widetilde{g}|\widehat{f}(\widetilde{\Theta}_m, \widetilde{k}_t)) + \frac{\widetilde{\eta}}{|\widetilde{g}|\widehat{f}(\widetilde{\Theta}_m, \widetilde{k}_t)} \end{aligned} \tag{36}$$

where (a) holds because $\ln(1+\epsilon) \approx \epsilon$. The relations in (36) indicate that the noise term will be magnified by $\frac{1}{|\widetilde{g}|\widehat{f}(\widetilde{\Theta}_m, \widetilde{k}_t)}$ times. Therefore, large errors will be introduced for small values of $\widehat{f}(\widetilde{\Theta}_m, \widetilde{k}_t)$. To avoid the noise amplification effects, we multiply (36) by $|\widetilde{y}_{m,t}|$ [14], and obtain

$$\gamma_{m,t} = |\widetilde{y}_{m,t}|\alpha + \frac{|\widetilde{y}_{m,t}|\widetilde{\Theta}_m}{\widehat{\sigma}_1^2}\Theta + \frac{|\widetilde{y}_{m,t}|\widetilde{k}_t}{\widehat{\sigma}_2^2}k + \widehat{\eta} \tag{37}$$

where

$$\begin{aligned}\gamma_{m,t} &= |\widetilde{y}_{m,t}|\ln|\widetilde{y}_{m,t}| + \frac{|\widetilde{y}_{m,t}|\widetilde{\Theta}_m^2}{2\widehat{\sigma}_1^2} + \frac{|\widetilde{y}_{m,t}|\widetilde{k}_t^2}{2\widehat{\sigma}_2^2} \\ \alpha &= \ln(|\widetilde{g}|\widehat{a}) - \frac{\Theta^2}{2\sigma_1^2} - \frac{k^2}{2\sigma_2^2}. \end{aligned} \tag{38}$$

Note that (37) is a linear equation of Θ and k. Similar to the procedures from (31) to (34), we can also obtain the estimation of Θ and k via the LS method. We omit the details and denote the estimation results as $\widehat{\Theta}_{\mathrm{WLS}}$ and $\widehat{k}_{\mathrm{WLS}}$. Then we can obtain $\widehat{r}_{\mathrm{WLS}}$ similar to (35).

Finally, we summarize the details of the proposed BRGA scheme in **Algorithm 1**.

Algorithm 1 Beam Refinement based on Gaussian Approximation (BRGA) Scheme

1: **Input:** N, N_{BS}, Q, S, M, $\widehat{q}$, $\widehat{s}$, $\widehat{a}$, $\widehat{\sigma}_1$ and $\widehat{\sigma}_2$.
2: Obtain $\widetilde{\boldsymbol{\Theta}}$ and $\widetilde{\boldsymbol{\mathcal{K}}}$ via (24) and (25).
3: Obtain $\widetilde{\Theta}_m$ and $\widetilde{k}_m$ for $m = 1, 2, \cdots, M$ via (26).
4: Obtain $\widetilde{y}_{m,t}$ via (27).
5: Obtain $\widehat{\Theta}_{\mathrm{LS}}$ and $\widehat{r}_{\mathrm{LS}}$ via (34) and (35), respectively.
6: Obtain $\widehat{\Theta}_{\mathrm{WLS}}$ and $\widehat{r}_{\mathrm{WLS}}$ similar to (34) and (35).
7: **Output:** $\widehat{\Theta}_{\mathrm{LS}}$, $\widehat{r}_{\mathrm{LS}}$, $\widehat{\Theta}_{\mathrm{WLS}}$ and $\widehat{r}_{\mathrm{WLS}}$.

Now, we evaluate the training overhead of the proposed BRGA scheme. As both $\widetilde{\boldsymbol{\Theta}}$ and $\widetilde{\boldsymbol{\mathcal{K}}}$ are quantized by M samples in (26), M^2 times of beam training are needed to obtain $\widetilde{y}_{m,t}$ in (27). Therefore, the training overhead of the proposed BRGA scheme is M^2.

IV. Simulation Results

Now we evaluate the performance of the proposed scheme. We consider an XL-MIMO system including a BS equipped with $N_{\mathrm{BS}} = 513$ antennas and a single-antenna user. We set the wavelength as $\lambda = 0.003$ m corresponding to the carrier frequency of 100 GHz. The channel between the BS and the user is composed of one LoS path and two NLoS paths, where the channel gain of the LoS path obeys $g_1 \sim \mathcal{CN}(0, 1)$ and that of the NLoS paths obeys $g_l \sim \mathcal{CN}(0, 0.01)$ for $l \in \{2, 3\}$. The channel angles distribute uniformly within $[-\sqrt{3}/2, \sqrt{3}/2]$, and the distance between the BS and the user distributes uniformly within $[10, 30]$ m. In addition, we set $Q = 513$ and $S = 6$ for the design of the codebook $\boldsymbol{C}$.

In Fig. 3, we compare the proposed BRGA scheme with the beam training only (BTO), ABP [8], and FFBR [10] in terms of the positioning error. We set $M = 2$ for the beam refinement, which corresponds to four times of additional channel tests. The deviation between the real position and the estimated position is denoted as E. From the figure, the BTO method performs the worst among all the methods. This is because the positioning results of the BTO are limited by the quantization deviation of the codebook. In addition, the ABP

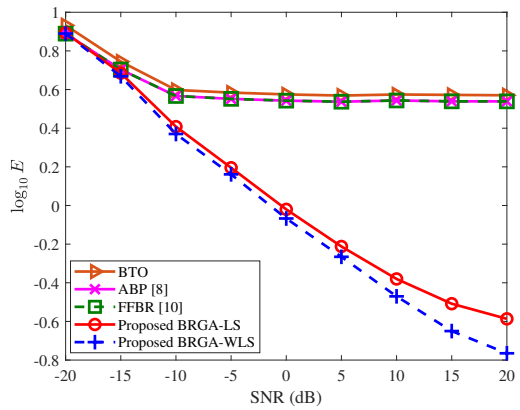

Fig. 3. Comparisons of the positioning error for different methods.

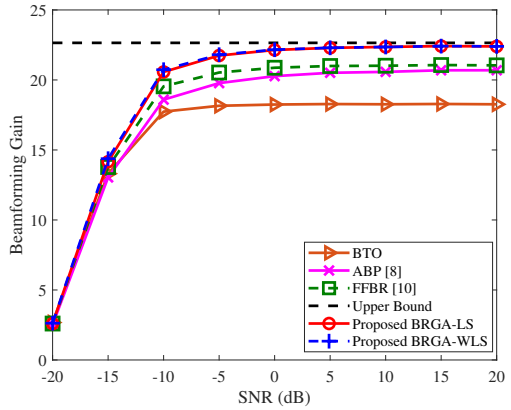

Fig. 4. Comparisons of the beamforming gains for different methods.

and FFBR have similar performance and slightly outperform the BTO benefiting from the refinement of the channel angles. Moreover, the proposed BRGA-LS and BRGA-WLS perform much better than the other methods due to the refinement of both the distance and channel angles, which validates the accuracy of the Gaussian approximation and verifies the effectiveness of the proposed BRGA scheme. It is also worth noting that the BRGA-WLS achieves better positioning accuracy than the BRGA-LS thanks to the elimination of the noise amplification effects.

In Fig. 4, we compare the proposed BRGA scheme with the BTO, ABP [8], and FFBR [10] in terms of the beamforming gain. The performances of the methods are consistent with those in Fig. 3 except that BRGA-LS and BRGA-WLS have similar performances while ABP and FFBR have different performances. This is because the positioning performance of BRGA-LS is accurate enough, and continuing to improve the positioning accuracy has little effect on the beamforming gain. On the other hand, the positioning performances of the ABP and FFBR are not accurate enough due to the lack of refinement for the distance, and the improvement of FFBR on ABP cannot be reflected in the positioning performance. In addition, both the BRGA-LS and the BRGA-WLS can approach the upper bound when SNR is larger than 0 dB, which verifies again the effectiveness of the BRGA scheme.

V. Conclusion

In this paper, we have investigated beam refinement for XL-MIMO. The BRGA scheme has been proposed and two estimators including BRGA-LS and BRGA-WLS have been developed to obtain high-accuracy channel parameter estimation. Simulation results have verified the effectiveness of the proposed scheme. Future work will be continued with the focus on effective beam training for XL-MIMO.

VI. Acknowledgment

This work is supported in part by National Natural Science Foundation of China (NSFC) under Grant 62071116 and U22B2007, and by National Key Research and Development Program of China under Grant 2021YFB2900404. The work of Kangjian Chen is supported in part by the Postgraduate Research&Practice Innovation Program of Jiangsu Province under Grant KYCX23_0262.

References

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

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

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

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

[5] 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.

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

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

[9] C. Qin, J. A. Zhang, X. Huang, K. Wu, and Y. J. Guo, "Fast angle-of-arrival estimation via virtual subarrays in analog antenna array," *IEEE Trans. Wireless Commun.*, vol. 19, no. 10, pp. 6425–6439, Oct. 2020.

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

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

[12] 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.

[13] T. F. Coleman and Y. Li, "An interior trust region approach for nonlinear minimization subject to bounds," *SIAM J. Optim.*, vol. 6, no. 2, pp. 418–445, 1996.

[14] K. Wu, J. A. Zhang, and Y. J. Guo, "Fast and accurate linear fitting for an incompletely sampled Gaussian function with a long tail [DSP Tips and Tricks]," *IEEE Signal Process. Mag.*, vol. 39, no. 6, pp. 76–84, Nov. 2022.