

Parallax-aware and MAB-empowered Beam Training for Modular XL-MIMO Systems

Chuanpeng Xu

Nanjing University of Posts and Telecommunications

Huibin Liang

Nanjing University of Posts and Telecommunications

Yunchao Song

songyc@njupt.edu.cn

Nanjing University of Posts and Telecommunications

Chen Liu

Nanjing University of Posts and Telecommunications

Yusong Wang

Nanjing University of Posts and Telecommunications

Research Article

Keywords:

Posted Date: October 27th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11265607/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Parallax-aware and MAB-empowered Beam Training for Modular XL-MIMO Systems

Chuanpeng Xu¹, Huibin Liang¹, Yunchao Song^{1*}, Chen Liu¹,
Yusong Wang¹

^{1*}College of Electronic and Optical Engineering & College of Flexible Electronics (Future Technology), Nanjing University of Posts and Telecommunications, Nanjing, Nanjing, 210023, Jiangsu, China.

*Corresponding author(s). E-mail(s): songyc@njupt.edu.cn;
Contributing authors: 1024020616@njupt.edu.cn;
2021020305@njupt.edu.cn; liuch@njupt.edu.cn;
2023020405@njupt.edu.cn;

Abstract

This paper proposes a parallax-aware (PA) and multi-armed bandit (MAB)-empowered beam-training scheme for modular extremely large-scale antenna arrays operating in hybrid near- and far-field environments. The scheme features a closed-loop synergy between adaptive codebook update and MAB learning. First, a PA initial codebook is constructed to combat near-field spatial non-stationarity. During probing, a greedy exponential Thompson sampling (G-Exp-TS) strategy guides beam selection, while an empirical quantile-quantile correlation criterion detects statistical steady states. Upon convergence, cross-subarray geometric projection based on spatial-frequency gradients is triggered, and a likelihood ratio test prunes identified noise-only sectors, achieving dynamic codebook compression. Within this compressed action space, G-Exp-TS optimally balances exploration and exploitation by casting beam selection as a 0-1 integer program to maximize long-term effective spectral efficiency (ESE). Simulation results demonstrate that the proposed scheme attains substantial ESE gains with markedly reduced training overhead.

Keywords: Modular extremely large antenna array, Hybrid-field, Beam training, Multi-armed bandit

1 Introduction

To meet the stringent data rate and capacity requirements of future 6G networks, wireless communications are inevitably migrating toward higher frequency bands, such as the millimeter wave and terahertz regimes [1, 2]. While abundant unallocated bandwidth is available, severe free-space path loss poses a fundamental challenge [3, 4]. Extremely large-scale multiple-input multiple-output (XL-MIMO) has emerged as a key physical-layer enabler to compensate for this propagation attenuation, where arrays comprising hundreds to thousands of antenna elements generate highly directional beams with massive beamforming gain [5, 6].

The large array apertures in XL-MIMO systems significantly extend the Rayleigh distance, causing a substantial number of users and scatterers to fall within the hybrid-field region, where near- and far-field electromagnetics coexist. To reduce the prohibitive beam training overhead in this mixed propagation regime, early studies employed hierarchical solutions that coarsely identify angles before refining near-field distances [7–9]. Later works addressed more complex phenomena, such as the hybrid-field beam squint effect [10], and leveraged deep learning to capture spatial frequency correlations [11]. A unified wavenumber-domain framework has also been proposed recently, reinterpreting spherical waves as plane-wave superpositions to simplify codebook design and curb probing complexity [12].

Conventional centralized XL-MIMO beam training heavily relies on full dimensional instantaneous channel state information (CSI), which incurs prohibitive pilot overhead and is highly vulnerable to transient channel fluctuations [13]. To further circumvent the pilot overhead of full-dimensional instantaneous CSI, statistical CSI (S-CSI) and spatial geometric priors have been exploited as pragmatic alternatives, characterizing the slowly varying layout of scattering clusters to constrain the beam search space [13, 14]. Accordingly, a growing number of recent works have utilized S-CSI for beam training in centralized XL-MIMO systems. To effectively leverage S-CSI, some beam training schemes focus on optimizing the codebook structure, including hierarchical spatial chirp codebooks [15], accumulation-based angular-radial decoupling [16], and wavenumber-domain transformation [12]. Furthermore, data-driven approaches have been widely adopted to learn the mapping from statistical features to beam indices. For instance, deep residual networks have been utilized to directly predict the optimal near field beam index from highly sparse pilot observations [17], while incorporating user position as a spatial prior has enabled adaptive adjustments of probing solutions [18].

While statistical priors alleviate the computational burden of beam training algorithms, the physical realization of fully connected architectures in massive scale systems remains severely hindered by staggering hardware complexity. Specifically, connecting every radio frequency (RF) chain to antennas requires an excessive number of phase-shifters, which leads to prohibitive power consumption. To overcome this fundamental hardware bottleneck, partially-connected subarray architectures have been widely adopted in conventional massive MIMO (M-MIMO) systems. For instance, the research demonstrates that a flexible unsupervised learning framework can effectively optimize subarray structures. This approach successfully accommodates practical constraints, such as noisy CSI and quantized phase-shifters [19]. As the antenna count

scales drastically, extending this modularity from M-MIMO systems to XL-MIMO paradigms yields even more pronounced advantages. Based on this architecture, initial research efforts focused on simplifying near field modeling within individual subarrays. For instance, computational complexity can be significantly reduced by applying far field approximations at the subarray level, supplemented by an analytical phase refinement utilizing the principle of stationary phase [20]. Furthermore, researchers recognized that the subarray architecture naturally accommodates spatial variations across massive apertures. To address spatial non-stationarity, [21] assigns distinct codewords to different subarrays to adaptively match localized channel conditions. Similarly, independent subarray beamforming has been exploited to capture and utilize ground-reflection multipath components in specific line-of-sight scenarios [22]. To further accelerate the search process, dynamically adjusting subarray activation topologies has become a prevalent strategy. Specifically, [23] utilizes a sparse Discrete Fourier Transform (DFT) codebook for initial coarse sweeping and subsequently activates the central subarray to resolve angular ambiguity. Despite these advances, existing subarray beam training methods rely on pre-defined codebooks or activation patterns, which limits their adaptability to dynamic, spatially non-stationary channels. To overcome this rigidity, the multi-armed bandit (MAB) framework provides a natural online learning alternative.

Rooted in reinforcement learning, the MAB model provides a robust framework for sequential decision-making under partial feedback. By iteratively probing candidate channels and leveraging accumulated channel statistics, MAB effectively balances the exploration-exploitation tradeoff. These characteristics make MAB particularly well-suited for beam training applications. In early beam training studies, MAB was primarily utilized to directly extract statistical priors from historical feedback. For instance, joint near-field angular and distance sweeping has been formulated as an MAB problem, employing the upper confidence bound (UCB) strategy to effectively identify high-gain beams [24]. To accelerate algorithmic convergence, researchers began incorporating environmental features and user collaborative information into the learning process. Specifically, contextual MAB algorithms have been introduced to bypass explicit CSI estimation [25], while collaborative learning policies drastically shorten beam training time by clustering users with similar signal-to-noise ratio preferences [26]. As system complexity scales, extending MAB to handle higher-dimensional action spaces has become a new focal point. To this end, hierarchical MAB architectures have been designed to jointly execute beamforming and user scheduling [27], and quantum walk principles have been incorporated to enhance exploration efficiency across 2D beam spaces [28]. Despite recent algorithmic advancements, directly applying existing MAB frameworks to XL-MIMO systems remains severely bottlenecked by the curse of dimensionality. Lacking dynamic action space compression, conventional MAB designs rely on static codebooks, inevitably suffering from prolonged blind exploration and excessively slow convergence.

For the XL-MIMO system in a hybrid-field environment, we propose a parallax-aware (PA) and MAB-empowered beam-training scheme. The key innovation of this scheme lies in the collaborative design of the parallax awareness and the online learning process, enabling dynamic compression of the codebook space without the need for

explicit CSI. Specifically, this scheme jointly operates the adaptive codebook update mechanism and the MAB strategy: using geometric priors to define the candidate set and using the statistical feedback to prune the codebook in real time, thereby achieving efficient beam training within an optimized search space. This mechanism avoids redundant exploration in pure noise sectors and significantly reduces the pilot overhead. The main contributions of this paper are summarized as follows:

- We propose a closed-loop synergy framework that fundamentally integrates adaptive codebook updating with MAB-empowered learning. Rather than treating the action space as static, this scheme utilizes the adaptive codebook update mechanism to convert the geometric priors of the modular array into dynamic candidate search regions. Subsequently, within this space, the MAB strategy is employed to extract statistical features from historical interactions for online decision-making. This collaborative approach avoids the high cost of global codebook scanning, significantly reduces the pilot overhead, and effectively improves the effective spectral efficiency (ESE).
- To overcome the severe parallax effect induced by spatial non-stationarity in hybrid-field environments, we derive a rigorous mathematical definition for beam index redundancy based on physical subarray displacement and the uniform power distance, thereby constructing an initial codebook that guarantees full spatial coverage. To enable dynamic codebook updating, we introduce a steady state detection mechanism based on empirical quantile-quantile (Q-Q) correlation. Once a steady state is reached, the system triggers cross-subarray geometric projection, coupled with a likelihood ratio (LR) test for zero-power detection. This dynamic update mechanism prunes invalid pure noise sectors in real time, achieving substantial codebook dimensionality reduction.
- To maximize the long-term ESE of the system, we formulate the beam training process as a MAB problem and equivalently transform the optimization objective into a 0-1 integer programming problem. To efficiently solve this, we propose a G-Exp-TS strategy. By exploiting the exponential posterior distribution of the received signal power, this strategy optimally balances the exploration of the unknown environment with the exploitation of known high-gain beams, thereby maximizing the long-term ESE.

2 System Model

2.1 Communication Scenario and Channel Model

We consider a frequency division duplexing XL-MIMO downlink communication system, where a base station (BS) serves a single antenna user. The BS is equipped with a M-ELAA comprising N modules, with each module containing M antennas and connected to a dedicated RF chain (i.e., the number of RF chains $N_{\text{RF}} = N$). The antennas are arranged in a uniform linear array (ULA) with a module aperture of $D_{\text{mod}} = (M - 1)d$, where $d = \lambda_c/2$ is the antenna spacing and λ_c is the carrier wavelength.

As illustrated in Fig. 1, the M-ELAA is deployed along the y -axis and is symmetric about the origin. Consequently, the coordinates of the m -th antenna in the n -th module ($n \in \{1, \dots, N\}$, $m \in \{1, \dots, M\}$) can be expressed as $\mathbf{p}_{n,m} = [0, \delta_{n,m}d]^T$, where $\delta_{n,m} = \frac{(2n-N-1)M+2m-M-1}{2}$.

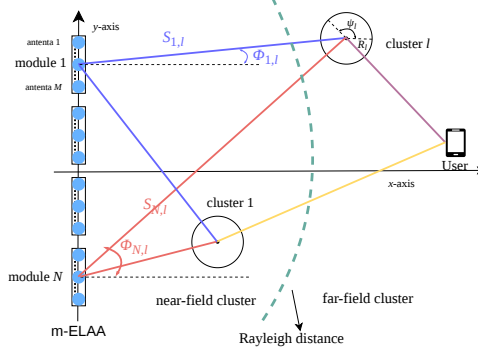


Fig. 1: Illustration of a multipath channel model.

In XL-MIMO systems, the large array aperture significantly extends the Rayleigh distance, defined as $Z = \frac{2((M-1)Nd)^2}{\lambda_c}$. Because this boundary spans a vast area, practical communication environments typically encompass scatterers located concurrently in both the near field and far field regions. This mixed spatial distribution gives rise to a hybrid-field propagation mechanism. Statistically, assuming correlated Rayleigh fading, this composite hybrid-field channel is characterized by

$$\mathbf{h}_t \sim \mathcal{CN}(\mathbf{0}, \mathbf{\Omega}), \quad (1)$$

where $\mathbf{\Omega}$ denotes the global spatial correlation matrix (i.e., the channel covariance matrix) spanning the entire massive array aperture. To accurately capture this hybrid field feature, we model the environment as comprising a total of L scattering clusters. Based on their physical distances to the array, these clusters are spatially partitioned into a near field set $\mathcal{L}_{\text{NF}}$ and a far field set $\mathcal{L}_{\text{FF}}$, satisfying $|\mathcal{L}_{\text{NF}}| + |\mathcal{L}_{\text{FF}}| = L$. Note that all scatterers are assumed to be located beyond the uniform power distance, i.e., within the uniform power region. The global spatial correlation matrix $\mathbf{\Omega} \in \mathbb{C}^{NM \times NM}$ is thus mathematically defined as the power weighted superposition of the components from all L clusters

$$\mathbf{\Omega} = \sum_{l \in \mathcal{L}_{\text{NF}}} \rho_l \mathbf{\Omega}_l^{\text{NF}} + \sum_{l \in \mathcal{L}_{\text{FF}}} \rho_l \mathbf{\Omega}_l^{\text{FF}}, \quad (2)$$

where ρ_l denotes the normalized average power fraction of the l -th cluster, satisfying $\sum_{l=1}^L \rho_l = 1$. For a near field cluster ($l \in \mathcal{L}_{\text{NF}}$), exact spherical wave propagation can be considered. The spatial correlation matrix $\mathbf{\Omega}_l^{\text{NF}}$ between the u -th and v -th antennas

across the entire M-ELAA is modeled as

$$[\mathbf{\Omega}_l^{\text{NF}}]_{u,v} = \int_{-\pi}^{\pi} e^{-j\frac{2\pi}{\lambda_c}(r_u(\psi_l)-r_v(\psi_l))} f(\psi_l) d\psi_l, \quad (3)$$

where $r_u(\psi_l)$ and $r_v(\psi_l)$ denote the exact physical distances from a scatterer within the l -th cluster to the u -th and v -th antennas, respectively. Specifically, the distance to the u -th antenna is given by $r_u(\psi_l) = \sqrt{r^2(\psi_l) + (\delta_u d)^2 - 2\delta_u d r(\psi_l) \sin(\theta_l(\psi_l))}$, where $\delta_u d$ represents the physical coordinate of the u -th antenna along the y -axis, and $r(\psi_l)$ is the distance from the scatterer to the array center. Furthermore, $\psi_l \in [-\pi, \pi]$ defines the angular spread within the l -th cluster, and $\theta_l(\psi_l)$ represents the global Angle of Arrival (AoA).

Conversely, for a far field cluster ($l \in \mathcal{L}_{\text{FF}}$), the impinging electromagnetic waves can be approximated as Uniform Plane Waves (UPW) exhibiting uniform power density and linear phase. The spatial correlation $\mathbf{\Omega}_l^{\text{FF}}$ simplifies to

$$[\mathbf{\Omega}_l^{\text{FF}}]_{u,v} = \int_{-\pi}^{\pi} e^{-j\frac{2\pi}{\lambda_c}(u-v)d\frac{S_l \sin \phi_l + R_l \sin \psi_l}{r(\psi_l)}} f(\psi_l) d\psi_l. \quad (4)$$

where $r(\psi_l)$ denote the physical distances from a scatterer within the l -th cluster to the array center, R_l represents the radius of the l -th cluster. In both cases, $f(\psi_l)$ characterizes the spatial power distribution, S_l and ϕ_l specify the physical central distance and angle of the l -th cluster relative to the array center. Specifically, it is defined as the power location spectrum for near-field clusters and the power angular spectrum for far-field clusters, both of which are assumed to follow a von-Mises distribution [29].

Although the aforementioned hybrid field model globally captures the spatial channel characteristics, the localized physical structure of the M-ELAA provides a highly practical simplification paradigm. It is crucial to note that the local Rayleigh distance for an individual module is profoundly reduced to $Z_{\text{mod}} = \frac{2((M-1)d)^2}{\lambda_c}$. Thus, the far-field approximation condition $Z_{\text{mod}} \ll r(\psi_l)$ is naturally satisfied for the scatterers, including those located within the global near field region. Accordingly, by applying the UPW approximation locally within the aperture of each subarray, the module specific spatial correlation matrix $\mathbf{\Omega}_n \in \mathbb{C}^{M \times M}$ can be analytically decoupled from the high dimensional global spherical wavefront, and it comprehensively aggregates the power from all L clusters

$$[\mathbf{\Omega}_n]_{m,m'} = \sum_{l=1}^L \rho_l \int_{-\pi}^{\pi} e^{-j\frac{2\pi}{\lambda_c}(m-m')d\frac{S_{n,l} \sin \phi_{n,l} + R_l \sin \psi_l}{r_n(\psi_l)}} f(\psi_l) d\psi_l, \quad (5)$$

where $m, m' \in \{1, \dots, M\}$ denote the intra module antenna indices, while $S_{n,l}$ and $\phi_{n,l}$ specify the physical distance and angle from the geometric center of the l -th cluster to the center of the n -th module. Meanwhile, $r_n(\psi_l)$ represents the exact distance from a scatterer to the n -th module.

Consequently, the global hybrid-field channel exhibits statistical decoupling at the subarray level, enabling the formulation of the channel vector for the n -th subarray as

$$\mathbf{h}_{n,t} \sim \mathcal{CN}(\mathbf{0}, \mathbf{\Omega}_n). \quad (6)$$

In coherence block t , let $\mathbf{h}_{n,t} \in \mathbb{C}^{M \times 1}$ denote the channel vector between the BS and the user for the n -th subarray. Accordingly, the signal received at the user from the n -th subarray is modeled as

$$y_{n,t} = \mathbf{h}_{n,t}^H \mathbf{w}_{n,t} x_{n,t} + \eta_{n,t}, \quad (7)$$

where $\mathbf{w}_{n,t} \in \mathcal{B}$ represents the active beamforming vector selected from the far-field DFT codebook for the n -th module. The term $x_{n,t}$ denotes the pilot symbol transmitted via the n -th RF chain, satisfying $\mathbb{E}[|x_{n,t}|^2] = 1$. Furthermore, $\eta_{n,t} \sim \mathcal{CN}(0, \sigma^2)$ characterizes the additive white Gaussian noise (AWGN).

2.2 Problem Formulation

Based on the local UPW property established earlier, the beamforming vector $\mathbf{w}_{n,t}$ for the n -th subarray is selected from a far-field DFT codebook $\mathcal{B} = \{\mathbf{b}_1, \dots, \mathbf{b}_M\}$ [7]. By aggregating the beamforming vectors across all N subarrays, the global pre-beamforming matrix is defined as $\mathbf{W}_t = \text{blkdiag}(\mathbf{w}_{1,t}, \mathbf{w}_{2,t}, \dots, \mathbf{w}_{N,t}) \in \mathbb{C}^{NM \times N}$. Consequently, the SE E_t achieved at coherence block t is calculated as

$$E_t = \log_2 \left(1 + \frac{\|\mathbf{h}_t^H \mathbf{W}_t\|_2^2}{\sigma^2} \right). \quad (8)$$

To account for the signaling costs in hybrid field environments, the ESE is further defined as

$$E_t^{\text{eff}} = \left(1 - \frac{T_{\text{BT}}}{T_{\text{sum}}} \right) E_t, \quad (9)$$

where T_{BT} and T_{sum} denote the beam training overhead and the total number of symbols per coherence block, respectively.

Assuming that the channel statistics remain stationary over T coherence blocks, the objective is to maximize the long term average ESE across all subarrays. The optimization problem $\mathcal{P}$ is formulated as follows

$$\begin{aligned} \mathcal{P} : \quad & \max_{\mathbf{w}_{n,t}} \frac{1}{T} \sum_{t=1}^T E_t^{\text{eff}} \\ \text{s.t.} \quad & \mathbf{w}_{n,t} \in \mathcal{B}, \quad \forall n. \end{aligned} \quad (10)$$

3 The PA And MAB-empowered Beam-training

The proposed PA and MAB-empowered beam training scheme comprises two collaborative components: an adaptive codebook update mechanism and MAB-based online

learning. As illustrated in Fig. 2, each coherence block is partitioned into codebook update, beam training and data transmission. Within this process, the MAB algorithm executes online beam training based on the currently codebook. Subsequently, by extracting statistical feedback from the training process, the adaptive codebook update mechanism dynamically curtails and prunes the redundant action space. By synergizing online learning with geometric pruning, the proposed scheme excludes redundant noise sectors and compresses training dimensionality, thereby reducing pilot overhead and accelerating beam convergence to maximize the ESE.

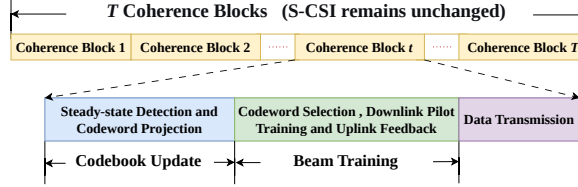


Fig. 2: The architecture of the PA and MAB-empowered beam-training scheme

3.1 Adaptive Codebook Update

This chapter presents the adaptive codebook update mechanism, comprising a PA initial codebook and its dynamic updating. It constructs the initial scanning space based on uniform power distance, and subsequently uses online statistical feedback to adaptively compress the search space. This mechanism filters redundant noise sectors to reduce candidate dimensionality, thereby saving resources, accelerating convergence, and boosting system ESE.

3.1.1 PA Initial Codebook Design

To reduce the high overhead of full codebook beam sweeping, this paper designs a low-overhead initial scanning codebook, denoted as $\mathcal{I}_{n,0}$. In XL-MIMO, users and scatterers mainly lie within the uniform power distance region. Thus, due to different subarray positions, their AoAs vary significantly, which is a near-field parallax effect. This severe parallax causes notable nonlinear shifts of optimal beam directions across M-ELAA modules. To address this, we leverage parallax boundaries to add precise angular overlap in adjacent scanning intervals. This PA-based initial codebook eliminates blind spots and ensures full spatial coverage.

Since the physical beamwidth varies non-linearly in the angular domain, we map the parallax effect into the spatial-frequency domain, defined as $\Theta = \sin \phi \in [-1, 1]$, which provides a uniform resolution. For a ULA comprising M antenna elements, the central spatial frequency of the i -th beam is typically expressed as $\Theta_i = \frac{2i-M-1}{M}$. In this domain, the resolution between any two adjacent beams remains constant at $\Delta\Theta_{\text{beam}} = \frac{2}{M}$.

To mathematically quantify the required angular redundancy between adjacent subarrays, let $D_{\min}$ denote the uniform power distance, and let D_{mod} denote the

physical spacing between adjacent subarrays. The required beam index redundancy, denoted by C_{red} , is established by the following lemma.

Lemma 1. *The beam index redundancy between adjacent sub-arrays is given by $C_{\text{red}} = \left\lceil \frac{MD_{\text{mod}}}{2D_{\text{min}}} \right\rceil$.*

Proof See Appendix A. □

Specifically, let $\mathcal{I}_{n,t}$ denote the active candidate beam index set allocated to the n -th subarray at any given coherence block t . To establish this dynamic action space, the entire codebook space is first uniformly partitioned, where the ideal base codeword index set allocated to the n -th subarray spans the range from $(n-1)\frac{M}{N} + 1$ to $n\frac{M}{N}$. Subsequently, to address the near field parallax, the initial candidate index set $\mathcal{I}_{n,0}$ at the initial coherence block ($t=0$) can be robustly constructed. Based on the derived required redundancy C_{red} , $\mathcal{I}_{n,0}$ can be formulated as

$$\mathcal{I}_{n,0} = \left\{ i \in \mathbb{Z}^+ \mid \max\left(1, (n-1)\frac{M}{N} + 1 - C_{\text{red}}\right) \leq i \leq \min\left(M, n\frac{M}{N} + C_{\text{red}}\right) \right\}. \quad (11)$$

Consequently, the initial codebook for the n -th subarray is formulated as

$$\mathcal{W}_{n,0} = \{\mathbf{b}_i \in \mathcal{B} \mid i \in \mathcal{I}_{n,0}\}, \quad (12)$$

thereby providing a highly reliable prior starting codebook for the subsequent high precision beam training.

3.1.2 Steady-state Detection and Codebook Update

Although the initial angular partitioning effectively reduces the search space, it inevitably assigns some subarrays to blind sectors with no users or scattering clusters. Fortunately, the high compactness of the initial codebook enables fast statistical convergence during the preliminary exploration phase, allowing the system to quickly identify subarrays that have captured high-energy beams. Once a steady state is detected, we exploit the spatial information shared across the array to geometrically project these validated beams onto other subarrays, triggering a dynamic cross-subarray codebook update that forces the entire array to focus solely on high-benefit directions.

Steady-state Detection: To precisely identify this statistical steady state and trigger the aforementioned projection at the optimal moment, we introduce a stability criterion based on empirical Q-Q correlation. Let $E_{n,\text{cur}}$ and $E_{n,\text{past}}$ denote the instantaneous SE sequences of the n -th subarray over the current window $\tau \in [t - L_{\text{wd}}, t - 1]$ and historical window $\tau \in [t - L_{\text{wd}} - L_{\text{step}}, t - L_{\text{step}} - 1]$, respectively. Let $\tilde{E}_{n,\text{cur}}$ and $\tilde{E}_{n,\text{past}}$ be their corresponding ascending sorted empirical quantiles. The Q-Q Pearson

correlation coefficient ρ_n is formulated as [30]

$$\rho_n = \frac{\sum_{l=1}^{L_{\text{wd}}} (\tilde{E}_{n,\text{cur}}^{(l)} - \bar{E}_{n,\text{cur}})(\tilde{E}_{n,\text{past}}^{(l)} - \bar{E}_{n,\text{past}})}{\sqrt{\sum_{l=1}^{L_{\text{wd}}} (\tilde{E}_{n,\text{cur}}^{(l)} - \bar{E}_{n,\text{cur}})^2} \sqrt{\sum_{l=1}^{L_{\text{wd}}} (\tilde{E}_{n,\text{past}}^{(l)} - \bar{E}_{n,\text{past}})^2}}, \quad (13)$$

where $\tilde{E}_{n,\text{cur}}^{(l)}$ and $\tilde{E}_{n,\text{past}}^{(l)}$ represent the l -th elements within their respective ascending sorted windows. Additionally, $\bar{E}_{n,\text{cur}}$ and $\bar{E}_{n,\text{past}}$ denote the sample means of the respective windows.

The training process is deemed to have reached a quasi-steady state if the correlation satisfies

$$\rho_n \geq \epsilon_{\text{corr}}, \quad \forall n \in \{1, 2, \dots, N\}, \quad (14)$$

where ϵ_{corr} is a rigorous tolerance threshold designed to strictly compensate for the artificial auto-correlation introduced by the window overlap.

Codebook Update: Once this stability threshold is reached, the system triggers the cross-subarray geometric projection. To execute this, the system leverages the statistical reward estimates collected during the preliminary exploration phase. Let $\hat{\mu}_{n,i,t}$ denote the empirical mean of the received signal power for the i -th beam at the n -th subarray up to coherence block t . To filter out weak multipath components and noise, we identify the global valid set $\mathcal{W}_t^{\text{can}}$, which aggregates all candidate index pairs (n, i) satisfying the pruning condition

$$\hat{\mu}_{n,i,t} \geq \alpha_{\text{th}} \hat{\mu}_{t,\text{max}}, \quad (15)$$

where $\hat{\mu}_{t,\text{max}} = \max_{n,i} \{\hat{\mu}_{n,i,t}\}$ represents the global maximum average beam gain recorded across the entire array at coherence block t , and $\alpha_{\text{th}} \in (0, 1)$ is a predefined relative threshold.

To project these validated beams across the array with low computational complexity, we exploit the spatial-frequency gradient. Let $\phi_{\bar{n},i^{\text{opt}}} = \arcsin((2i^{\text{opt}} - M - 1)/M)$ be the physical angle of a validated high-gain beam i^{opt} at the source subarray $\bar{n}$. When projecting this beam to a target subarray n separated by a physical displacement $D_{\bar{n},n}$, the far-field extreme assumes a planar wavefront with zero parallax shift, inherently keeping the projected beam index identical to the source

$$b_{\text{far}} = i^{\text{opt}}. \quad (16)$$

In contrast, the near field extreme is governed by the spherical wavefront effect, which induces a noticeable parallax shift across different subarrays. This shift reaches its maximum when the signal source or scatterer is located at the minimum uniform power distance, D_{min} . To analytically quantify this worst-case spatial deviation, we apply a first-order spatial-frequency gradient approximation (similar to the proof of Lemma 1). Specifically, the maximum index drift from the source subarray $\bar{n}$ to the target subarray n is determined by

$$\Delta b_{\bar{n} \rightarrow n} = \frac{MD_{\bar{n},n} \cos^3 \phi_{\bar{n},i^{\text{opt}}}}{2D_{\text{min}}}. \quad (17)$$

Consequently, the near-field projected index extreme is bounded at

$$b_{\text{near}} = i^{\text{opt}} + \Delta b_{\bar{n} \rightarrow n}. \quad (18)$$

Note that the physical displacement $D_{\bar{n},n}$ is a signed value depending on the relative positions of the subarrays. Consequently, the near field limit b_{near} may shift to either the left or the right of the far field reference b_{far} . To account for this directional ambiguity and ensure a robust projection, the predicted candidate index interval $\mathcal{Q}_{\bar{n},i^{\text{opt}}}$ originating from the source beam i^{opt} is rigorously formed by spanning all discrete indices between these two spatial extremes

$$\mathcal{Q}_{\bar{n},i^{\text{opt}}} \triangleq \{ \lfloor \min(b_{\text{near}}, b_{\text{far}}) \rfloor, \dots, \lceil \max(b_{\text{near}}, b_{\text{far}}) \rceil \}. \quad (19)$$

The preliminary candidate index set $\tilde{\mathcal{I}}_{n,t}$ is then constructed by taking the union of its natively retained valid indices $\mathcal{I}_{n,t}^{\text{opt}}$ and the gradient-based geometric projections from all other subarrays

$$\tilde{\mathcal{I}}_{n,t} = \mathcal{I}_{n,t}^{\text{opt}} \cup \bigcup_{\substack{(\bar{n},i^{\text{opt}}) \in \mathcal{W}_t^{\text{can}} \\ \bar{n} \neq n}} \mathcal{Q}_{\bar{n},i^{\text{opt}}}. \quad (20)$$

However, the cross-subarray geometric projection inevitably introduces some low-gain or pure-noise codewords into the candidate set, leading to an unnecessary expansion of the codebook dimension. To address this, we employ an LR test for zero-power detection to dynamically prune the codebook.

We employ a LR test for zero-power detection. For any codeword $\mathbf{b}_i \in \mathcal{I}_{n,t}$ evaluated at coherence block t , the BS transmits a pilot symbol x_t satisfying $|x_t|^2 = 1$. The received signal is $y_{n,i,t} = \mathbf{h}_{n,t}^H \mathbf{b}_i x_t + \eta_{n,t}$, where $\eta_{n,t}$ denotes the AWGN. Multiplying $y_{n,i,t}$ by the conjugate pilot x_t^* yields the equivalent scalar observation

$$\tilde{y}_{n,i,t} = y_{n,i,t} x_t^* = \mathbf{h}_{n,t}^H \mathbf{b}_i + \tilde{\eta}_{n,t}, \quad (21)$$

where $\tilde{\eta}_{n,t} = \eta_{n,t} x_t^*$. Suppose the i -th beam has been selected $C_{n,i,t}$ times up to coherence block t . Let $\tilde{\mathbf{y}}_{n,i} = [\tilde{y}_{n,i,1}, \dots, \tilde{y}_{n,i,C_{n,i,t}}]$ denote the collection of historical observations. The LR is defined as

$$\gamma_{n,i,t} = \frac{\int_{\zeta_{\text{LR}}}^{\infty} \mathbb{P}(\tilde{\mathbf{y}}_{n,i} | p_{\text{LR}}) d p_{\text{LR}}}{\int_{\sigma^2}^{\zeta_{\text{LR}}} \mathbb{P}(\tilde{\mathbf{y}}_{n,i} | p_{\text{LR}}) d p_{\text{LR}}}, \quad (22)$$

where p_{LR} denotes the expected power and ζ_{LR} is a predefined power threshold. Distinctly, a large LR value ($\gamma_{n,i,t} > \epsilon_{\text{LR}}$) confidently indicates a valid high-gain beam, whereas a small value ($\gamma_{n,i,t} \leq 1/\epsilon_{\text{LR}}$) identifies a zero-power (pure-noise) beam. For the analytical closed-form calculation of $\gamma_{n,i,t}$, please refer to [31].

To adaptively compress the action space, let $\Upsilon_{n,t}$ denote the pruned set of beam indices for the n -th subarray. This set mathematically aggregates all statistically

proven zero-power beams whose LR values fall below the pruning threshold $1/\epsilon_{\text{LR}}$, i.e

$$\Upsilon_{n,t} = \left\{ i \in \mathcal{M} \mid \gamma_{n,i,t} \leq \frac{1}{\epsilon_{\text{LR}}} \right\}, \quad (23)$$

where $\mathcal{M} = \{1, 2, \dots, M\}$ represents the global codebook index space. On the other hand, beams yielding $\gamma_{n,i,t} > \epsilon_{\text{LR}}$ are successfully identified as reliable active candidates. To reduce the computational complexity, these confirmed high-gain beams are subsequently exempted from redundant statistical testing and are directly maintained in the candidate set. Consequently, the final effective candidate index set $\mathcal{I}_{n,t}$ used for the subsequent MAB exploration phase is obtained by removing these statistically proven noise beams from the preliminary geometrically projected set $\tilde{\mathcal{I}}_{n,t}$:

$$\mathcal{I}_{n,t} = \left\{ i \in \tilde{\mathcal{I}}_{n,t} \mid i \notin \Upsilon_{n,t} \right\}.$$

The dynamic codebook at coherence block t is strictly updated as

$$\mathcal{W}_{n,t} = \{\mathbf{b}_i \in \mathcal{B} \mid i \in \mathcal{I}_{n,t}\}. \quad (24)$$

3.2 MAB-empowered Beam Training

The MAB framework is a classic reinforcement learning paradigm that models sequential decision-making under uncertainty. In a standard MAB problem, an agent iteratively selects from a set of available actions, each yielding a reward drawn from an initially unknown probability distribution. The primary objective is to maximize the cumulative reward over a given time horizon. To achieve this, the agent carefully navigates the exploration-exploitation dilemma: balancing the exploration of less-sampled arms to learn their true reward distributions against the exploitation of the arm currently believed to be optimal.

Guided by this paradigm, we cast the dynamic beam training process as a MAB problem. To balance system performance with exploration speed, we denote the training overhead budget per subarray by G , and let $\mathcal{A}_{n,t}$ be the index set of selected codewords from the adaptively refined candidate codebook $\mathcal{W}_{n,t}$ in coherence block t . Given that the beam probing is implemented using the candidate codewords $\mathbf{b}_i$, $i \in \mathcal{I}_{n,t}$, the problem is to find $\mathcal{A}_{n,t}$ to maximize the total received energy across all subarrays, which is formulated as

$$\begin{aligned} \mathcal{P}_1 : \quad & \max_{\mathcal{A}_{n,t}} \frac{1}{T} \sum_{t=1}^T \sum_{n=1}^N \sum_{i \in \mathcal{A}_{n,t}} |\mathbf{h}_{n,t}^H \mathbf{b}_i|^2 \\ \text{s.t.} \quad & \mathcal{A}_{n,t} \subseteq \mathcal{I}_{n,t} \\ & |\mathcal{A}_{n,t}| = G. \end{aligned} \quad (25)$$

Since explicit CSI is practically unavailable, we formulate the beam training process as a MAB problem, which is inherently capable of learning statistical channel features

from online feedback. Specifically, each candidate codeword $\mathbf{b}_i \in \mathcal{W}_{n,t}$ is regarded as a base arm, and the selected subset $\mathcal{A}_{n,t}$ constitutes a super arm. The instantaneous reward for pulling the i -th base arm is defined as the received signal power

$$\mu_{n,i,t} = |\tilde{y}_{n,i,t}|^2 = |\mathbf{h}_{n,t}^H \mathbf{b}_i + \tilde{\eta}_{n,t}|^2. \quad (26)$$

To design a rigorous sampling strategy, the statistical distribution of this reward can be explicitly characterized.

Lemma 2. *Let y be a complex random variable following a zero-mean complex Gaussian distribution, $y \sim \mathcal{CN}(0, \varrho)$. The squared modulus $\mu = |y|^2$ follows an exponential distribution with mean ϱ , i.e., $\mu \sim \text{Exp}(1/\varrho)$.*

Proof See Appendix B. □

To effectively solve MAB problems, Thompson Sampling (TS) has emerged as a highly efficient Bayesian heuristic. Unlike the UCB strategies, TS constructs a posterior distribution for the expected reward of each arm and randomizes action selection by sampling directly from these posteriors. This probabilistic framework naturally calibrates the exploration-exploitation trade-off: arms with high expected rewards or high uncertainty are inherently more likely to be selected. In our hybrid-field beam training scenario, TS is particularly advantageous because it can explicitly leverage the exponential distribution of the signal power to construct precise posterior distributions, guaranteeing fast convergence and asymptotic optimality. To balance exploration and exploitation, we adopt the G-Exp-TS strategy. In coherence block t , for each arm $i \in \mathcal{W}_{n,t}$, G-Exp-TS generates a sampling value $\beta_{n,i,t}$ governed by an exploration probability $p_{\text{exp}} = \frac{1}{|\mathcal{W}_{n,t}|}$

1) Greedy Exploitation: With probability $1 - p_{\text{exp}}$, it directly assigns the empirical mean $\beta_{n,i,t} = \hat{\mu}_{n,i,t}$.

2) Posterior Exploration: With probability p_{exp} , the algorithm explores the arm by sampling from its exponential posterior distribution. Specifically, by defining $\vartheta_{n,i,t} = \frac{C_{n,i,t}-1}{C_{n,i,t}}$ (where $C_{n,i,t}$ is the selection count) and drawing a uniform random variable $\zeta \sim \mathcal{U}(0, 1)$, the exponential sampling value $\beta_{n,i,t}$ is obtained by solving the bounds derived from the Kullback-Leibler (KL) divergence. The KL divergence serves as a fundamental statistical metric, quantifying how the empirical probability distribution of the rewards diverges from the hypothesized true distribution. In this context, the KL divergence function is explicitly defined as

$$D_{\text{KL}}(\hat{\mu}_{n,i,t} \parallel \iota) = \ln \left(\frac{\iota}{\hat{\mu}_{n,i,t}} \right) + \frac{\hat{\mu}_{n,i,t}}{\iota} - 1. \quad (27)$$

If $\zeta \geq 0.5$, we solve for the right-tail target KL divergence bound

$$\beta_{n,i,t} = \sup \left\{ \iota \geq \hat{\mu}_{n,i,t} : D_{\text{KL}}(\hat{\mu}_{n,i,t} \parallel \iota) \leq \frac{-\ln(2(1-\zeta))}{C_{n,i,t} \cdot \vartheta_{n,i,t}} \right\}. \quad (28)$$

If $\zeta < 0.5$, we solve for the lower bound of the transcendental equation

$$\beta_{n,i,t} = \inf \left\{ \iota \leq \hat{\mu}_{n,i,t} : D_{\text{KL}}(\hat{\mu}_{n,i,t} \parallel \iota) \leq \frac{-\ln(2\zeta)}{C_{n,i,t} \cdot \vartheta_{n,i,t}} \right\}. \quad (29)$$

Subsequently, to cast the selection of the super arm $\mathcal{A}_{n,t}$ as a specific optimization problem, we revisit the optimization problem $\mathcal{P}_1$. Let $\mathbf{v}_{n,t} \in \{0, 1\}^{|\mathcal{W}_{n,t}|}$ act as the binary decision vector in coherence block t , where its i -th element $v_{n,i,t} = 1$ implies the i -th codeword is selected, and $v_{n,i,t} = 0$ otherwise. By utilizing the generated sampling value vector $\boldsymbol{\beta}_{n,t} = [\beta_{n,1,t}, \dots, \beta_{n,|\mathcal{W}_{n,t}|,t}]^T$ as a surrogate for the expected reward, problem $\mathcal{P}_1$ can be equivalently transformed into the following 0-1 integer programming problem

$$\begin{aligned} \mathcal{P}_2 : \quad & \max_{\mathbf{v}_{n,t}} \boldsymbol{\beta}_{n,t}^T \mathbf{v}_{n,t} \\ \text{s.t.} \quad & \mathbf{e}^T \mathbf{v}_{n,t} = G, \end{aligned} \quad (30)$$

where $\mathbf{e}$ is an all-ones column vector of length $|\mathcal{W}_{n,t}|$.

The optimal solution dictates selecting the top G candidate codewords from $\mathcal{I}_{n,t}$ that yield the highest sampling values. For each selected codeword $\mathbf{b}_i \in \mathcal{A}_{n,t}$, its corresponding selection count and empirical mean are updated in real-time as:

$$C_{n,i,t+1} = C_{n,i,t} + 1, \quad (31)$$

$$\hat{\mu}_{n,i,t+1} = \frac{\hat{\mu}_{n,i,t} C_{n,i,t} + \mu_{n,i,t}}{C_{n,i,t+1}}. \quad (32)$$

Theorem 1 *Over T coherence blocks, the expected total regret of the proposed scheme is bounded by $O(N\sqrt{V_E MT})$, where V_E is the variance bound.*

Proof See Appendix C. □

The complete procedure of the proposed geometry-guided dynamic beam training is summarized in Algorithm 1. By fundamentally limiting the MAB action space to $\mathcal{W}_{n,t}$, the algorithm avoids redundant explorations in noise sectors and focuses entirely on refining the optimal beams, thereby accelerating convergence and maximizing spectral efficiency.

4 Simulation Results

In this section, we evaluate the performance of the proposed scheme through numerical simulations. The primary system parameters are set as follows: $f_c = 100$ GHz, $N = 8$, $M = 32$, and $D_{\min} = 1.3$ m. The central angles of the scattering clusters are uniformly distributed in $[-\pi/3, \pi/3]$, and their distances are uniformly sampled from $\{4, 8, 12, 50, 70, 110, 120\}$ m. Furthermore, the beam training overhead budget G is tightly coupled with the number of scattering clusters, satisfying $G = L$. For the proposed algorithm, the parameters are configured as $L_{\text{wd}} = 20$, $L_{\text{step}} = 5$, $\epsilon_{\text{corr}} =$

Algorithm 1 PA and MAB-empowered Beam Training

Require: $\mathcal{B}$, G , L_{wd} , L_{step} , ϵ_{corr} , α_{th} , ϵ_{LR} , ζ_{LR} , σ^2

Ensure: $\mathcal{I}_{n,t}$, $\mathcal{W}_{n,t}$, and $\mathcal{A}_{n,t}$, $\forall n \in \{1, \dots, N\}$

```

1: Initialize  $C_{\text{red}} = \lceil \frac{MD_{\text{mod}}}{2D_{\text{min}}} \rceil$ ,  $\mathcal{I}_{n,0}$ ,  $\mathcal{W}_{n,0}$ ,  $\hat{\mu}_{n,i,0} = 0$ ,  $C_{n,i,0} = 0$ ,  $\Upsilon_{n,0} = \emptyset$ , and  $\tilde{\mathbf{y}}_{n,i} = \emptyset$ ;
2: for  $t = 1, \dots, T$  do
3:   if  $t > L_{\text{wd}} + L_{\text{step}}$  and  $\rho_n \geq \epsilon_{\text{corr}}, \forall n$  then
4:     Calculate  $\hat{\mu}_{t,\text{max}}$  and determine  $\mathcal{W}_t^{\text{can}}$ ;
5:     for  $n = 1, \dots, N$  do
6:       Compute  $\Delta b_{\bar{n} \rightarrow n}$  using (17);
7:       Calculate  $\mathcal{Q}_{\bar{n},i^{\text{opt}}}$  and obtain  $\tilde{\mathcal{I}}_{n,t}$ ;
8:       Update  $\mathcal{I}_{n,t}$  and  $\mathcal{W}_{n,t}$ ;
9:     end for
10:   else
11:      $\mathcal{I}_{n,t} = \mathcal{I}_{n,t-1}$  and  $\mathcal{W}_{n,t} = \mathcal{W}_{n,t-1}$ ;
12:   end if
13:   for  $n = 1, \dots, N$  do
14:     Generate  $\beta_{n,i,t}$  for  $i \in \mathcal{I}_{n,t}$  via G-Exp-TS;
15:     Solve  $\mathcal{P}_2$  to obtain  $\mathcal{A}_{n,t}$ ;
16:     for  $i \in \mathcal{A}_{n,t}$  do
17:       Acquire  $\tilde{y}_{n,i,t}$  and update  $\tilde{\mathbf{y}}_{n,i}$ ;
18:       Update  $C_{n,i,t}$  and  $\hat{\mu}_{n,i,t}$ ;
19:       Compute  $\gamma_{n,i,t}$ ;
20:     end for
21:     Update  $\Upsilon_{n,t}$ ;
22:   end for
23: end for

```

0.95, $\alpha_{\text{th}} = 0.2$, and $\epsilon_{\text{LR}} = 10$. Assuming a coherence time of 1.5 ms and a user terminal velocity of 1 m/s, the CCM remains quasi-static over a stationary period of $T = 667$ coherence blocks.

To validate the superiority of our proposed scheme (defined in the legend as G-Exp-TS BT), we evaluate its performance against the following beam training baselines:

- TH BT [20]: A scheme specifically designed to perform hybrid-field beam training for partially-connected subarray architectures.
- Dual-UCB BT [24]: A centralized beam training framework that leverages the MAB learning mechanism to optimize the beam training process.
- Hybrid BT [8]: A centralized scheme that differentiates between near-field and far-field users through coarse beam sweeping with partially activated antennas, subsequently employing a fine-grained hierarchical codebook to achieve precise beam alignment.

Complexity Analysis. For **Algorithm 1**, the per-coherence-block complexity is dominated by the dynamic codebook update and G-Exp-TS beam training. In the worst case, each subarray retains the full codebook of size M . The valid beam screening

and cross-subarray projection in Lines 3-8 require $O(NM)$ and $O(N^2M)$ operations, respectively. The G-Exp-TS sampling and $\mathcal{P}_2$ solution in Lines 14-15 require $O(NM)$ and $O(NGM)$ operations, respectively, while the update and LR test in Lines 16-21 cost $O(NG)$. Hence, the worst-case per-coherence-block complexity is $O(NM + N^2M + NM + NGM + NG) = O(NGM + N^2M)$.

The per-coherence-block complexities of the three baseline schemes are as follows: the TH BT scheme has a complexity of $O(M)$; the Dual-UCB BT scheme has a complexity of $O(N + M^f S_{\text{dis}})$, where M^f denotes the number of beams selected in the first stage, and S_{dis} is the number of distance samples; the Hybrid BT scheme incurs a training overhead of $O(\alpha N + 2 \log(N))$, with α the fraction of activated antennas.

Fig. 3 illustrates the ESE versus the number of coherence blocks under different numbers of scattering clusters ($L \in \{1, 2, 3\}$), where the SNR is set to 10 dB. As observed, the ESE initially increases during the exploration phase and then rapidly converges to a stable steady state. This is because the steady-state detection and cross-subarray projection effectively prune the noise-dominated candidate beams, significantly reducing the exploration space for the subsequent MAB process. Additionally, the steady-state ESE is highest when $L = 1$ due to the highly concentrated signal energy. Although the spatial energy dispersion slightly reduces the maximum ESE as L increases, the proposed algorithm consistently maintains fast convergence and robustness across all multipath conditions.

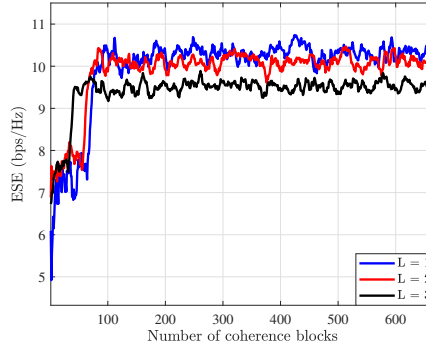


Fig. 3: EAE versus coherence blocks.

Fig. 4 compares the ESE performance of different beam training schemes versus SNR. Fig. 4(a) and Fig. 4(b) present the results for $L = 1$ and $L = 3$, respectively. It can be seen that the proposed G-Exp-TS BT scheme achieves a higher ESE than the TH BT, Dual-UCB BT, and Hybrid BT algorithms across the entire evaluated SNR range. By adaptively compressing the codebook and filtering out zero-power beams, the proposed scheme can efficiently focus on the optimal directions, thereby maximizing the system sum rate. Conversely, the baseline schemes exhibit inferior performance, they allocate a significant portion of the training overhead to exploring non-optimal or low-gain arms, which limits their ESE.

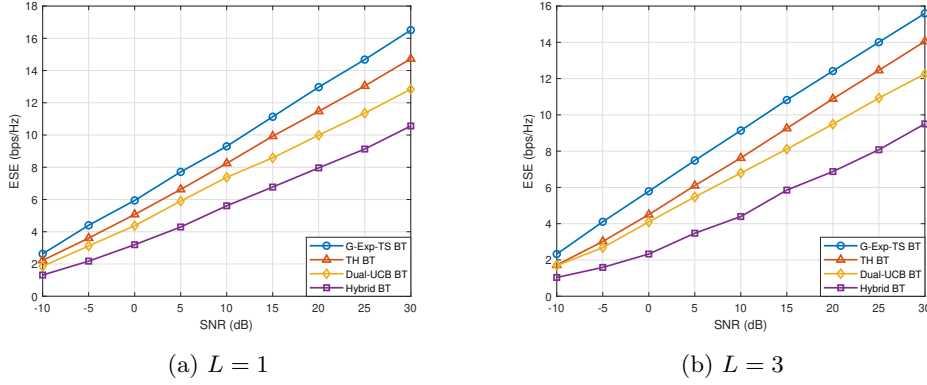
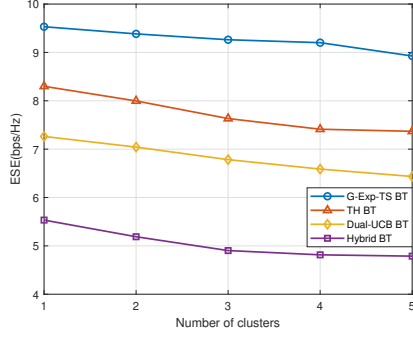


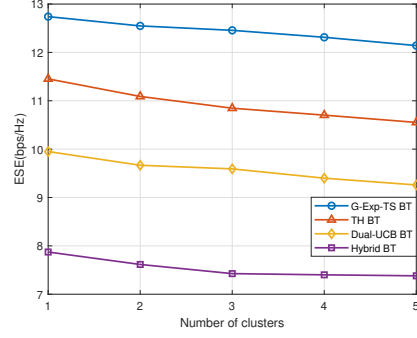
Fig. 4: ESE versus SNR for different numbers of scattering clusters.

Fig. 5 investigates the impact of the number of scattering clusters on the ESE under SNRs of 10 dB and 20 dB, as shown in Fig. 5(a) and (b), respectively. As observed in both scenarios, the ESE of all schemes exhibits a downward trend as the number of clusters L increases. This performance degradation occurs because the total channel energy becomes spatially dispersed across multiple multipath components, fundamentally limiting the maximum power that can be captured by a fixed number of RF chains. Although the higher SNR in Fig. 5(b) naturally elevates the overall ESE baseline compared to Fig. 5(a), the severe energy dispersion remains the primary bottleneck. Despite these challenging conditions, the proposed algorithm consistently outperforms the benchmark schemes and tightly tracks the performance upper bound. This robustness confirms that the dynamic geometric projection and G-Exp-TS exploration effectively identify and exploit the dominant hybrid-field paths, thereby minimizing the impact of severe multipath interference across different noise regimes.

Figs. 6(a) and (b) depict the ESE versus the beam training overhead budget under SNRs of 10 dB and 20 dB for $L = 3$, respectively. It can be observed that the Dual-UCB BT and Hybrid BT schemes exhibit a distinct inverted-U trend: although increasing the overhead budget initially facilitates convergence and enhances the ESE, their uncompressed codebooks inevitably lead to excessive exploration once the optimal performance-overhead balance is surpassed. This over-exploration incurs a severe time-fraction penalty, which subsequently degrades the ESE. In contrast, the proposed G-Exp-TS BT scheme achieves its peak ESE at a minimal overhead budget and consistently outperforms all baseline schemes. Benefiting from the geometry-guided spatial projection and the LR-based pruning mechanism, the dimensionality of the dynamic candidate codebook is drastically compressed. Consequently, even when a large overhead budget is allocated, the actual training overhead is strictly bounded by the small number of retained valid codewords. This mechanism fundamentally circumvents additional time-fraction penalties, enabling the system's ESE to plateau steadily. Meanwhile, since the overhead of the TH BT scheme is determined solely



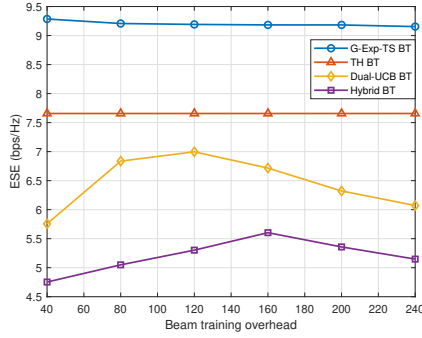
(a) SNR = 10 dB



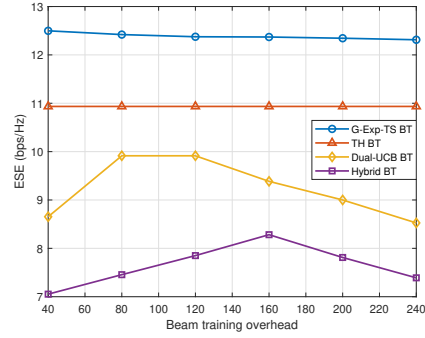
(b) SNR = 20 dB

Fig. 5: ESE versus the number of scattering clusters at different SNRs.

by the number of antennas per subarray, its ESE remains completely invariant to the dynamic overhead budget, manifesting as a horizontal line.



(a) SNR = 10 dB



(b) SNR = 20 dB

Fig. 6: ESE versus training overhead at different SNRs ($L = 3$).

5 Conclusion

In this paper, we proposed a PA beam training scheme for M-ELAA in hybrid-field environments. By integrating a modular antenna architecture with MAB online learning, the proposed framework features a powerful synergy between adaptive codebook updating and MAB learning to bypass the prohibitive overhead of traditional exhaustive sweeping. Specifically, to combat near field spatial non-stationarity, we first constructed a PA initial codebook. During the preliminary phase, a G-Exp-TS algorithm is employed to conduct beam probing over this initial space, while an

empirical Q-Q correlation criterion continuously monitors the process for statistical steady states. Upon detecting convergence, the system leverages the acquired statistical reward estimates to execute spatial-frequency gradient-based cross-subarray projections. By applying a LR test to strictly prune pure-noise sectors, the codebook was adaptively updated and drastically compressed. Subsequently, the G-Exp-TS algorithm successfully adapted to balance beam exploration and exploitation within this reduced codebook space. Simulation results demonstrates that the proposed scheme achieves high ESE with minimal overhead, significantly enhancing the effective spectral efficiency. Future work will investigate spatial non-stationarity induced by partial visibility in both single-user and multi-user scenarios to develop robust beam training methods for complex blockage environments.

Declarations

- **Funding:** This work was supported in part by the National Natural Science Foundation of China under Grant 62371249, and in part by the Key Project of Natural Science Foundation of Jiangsu Province (BE2023087).
- **Conflict of interest/Competing interests:** The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.
- **Ethics approval and consent to participate:** Not applicable.
- **Consent for publication:** Not applicable.
- **Data availability:** The data that support the findings of this study are available from the corresponding author upon reasonable request.
- **Materials availability:** Not applicable.
- **Code availability:** Not applicable.
- **Author contribution:** Chuanpeng Xu: Conceptualization, Methodology, Software, Writing – original draft. Huibin Liang: Data curation, Investigation, Validation. Yunchao Song: Supervision, Funding acquisition, Writing – review & editing. Chen Liu: Formal analysis, Visualization. Yusong Wang: Software, Validation. All authors read and approved the final manuscript.

Appendix

Proof of Lemma 1

To analytically quantify the parallax effect, let the center coordinates of the n -th subarray and the user be denoted by $(0, q_n)$ and (u, v) in a 2D Cartesian coordinate system, respectively. Owing to the uniform resolution characteristic of the spatial-frequency domain, the number of redundant beams required for seamless coverage can be precisely determined by quantifying the total observation drift of the near-field spatial frequency.

Let the spatial frequency $\Theta(q_n)$ be expressed as a function of the array position variable q_n , such that

$$\Theta(q_n) = \sin \phi_n = \frac{v - q_n}{\sqrt{u^2 + (v - q_n)^2}}. \quad (33)$$

The rate of change of the spatial frequency with respect to the continuous displacement of subarrays is governed by the partial derivative

$$\left| \frac{\partial \Theta(q_n)}{\partial q_n} \right| = \frac{u^2}{(u^2 + (v - q_n)^2)^{3/2}} = \frac{\cos^3 \phi_n}{u}, \quad (34)$$

where ϕ_n is the angle of arrival from the user relative to the broadside normal of module n .

For adjacent subarrays separated by a discrete physical distance D_{mod} , since the shortest perpendicular distance to the user is strictly bounded by $u \geq D_{\text{min}}$, the first-order spatial-frequency drift $\Delta\Theta$ at a specific observation angle ϕ_n is upper-bounded by

$$\Delta\Theta \approx \left| \frac{\partial \Theta(q_n)}{\partial q_n} \right| D_{\text{mod}} \leq \frac{D_{\text{mod}} \cos^3 \phi_n}{D_{\text{min}}}. \quad (35)$$

Furthermore, considering the worst-case geometric scenario across all possible directions where $\cos \phi_n \leq 1$, the global maximum spatial-frequency drift $\Delta\Theta_{\text{max}}$ is firmly bounded by

$$\Delta\Theta_{\text{max}} \leq \frac{D_{\text{mod}}}{D_{\text{min}}}. \quad (36)$$

Consequently, to thoroughly absorb this maximum global drift and ensure blind-spot-free coverage, the required beam index redundancy C_{red} is evaluated by dividing this drift by the constant beam resolution $\Delta\Theta_{\text{beam}} = 2/M$:

$$C_{\text{red}} = \left\lceil \frac{\Delta\Theta_{\text{max}}}{\Delta\Theta_{\text{beam}}} \right\rceil = \left\lceil \frac{MD_{\text{mod}}}{2D_{\text{min}}} \right\rceil. \quad (37)$$

This concludes the proof.

Proof of Lemma 2

From the channel model, the received baseband signal follows a zero-mean circularly symmetric complex Gaussian distribution, i.e., $y_{n,i,t} \sim \mathcal{CN}(0, \varrho)$, where the variance is defined as $\varrho = \mathbf{b}_i^H \mathbf{\Omega}_n \mathbf{b}_i + \sigma^2$. Let $X = \text{Re}(y_{n,i,t})$ and $Y = \text{Im}(y_{n,i,t})$. They are independent and identically distributed (i.i.d.) real Gaussian random variables, i.e., $X, Y \sim \mathcal{N}(0, \frac{\varrho}{2})$. The joint probability density function (PDF) of X and Y is given by

$$f_{X,Y}(x, y) = \frac{1}{\pi \varrho} \exp \left(-\frac{x^2 + y^2}{\varrho} \right). \quad (38)$$

Let $Z = |y_{n,i,t}|^2 = X^2 + Y^2$ denote the squared modulus of the signal. The cumulative distribution function (CDF) of Z can be evaluated by integrating the joint PDF over the circular region $x^2 + y^2 \leq z$. By transforming the Cartesian coordinates into polar coordinates $(x = r \cos \theta, y = r \sin \theta)$, the CDF is derived as

$$F_Z(z) = \int_0^{2\pi} \int_0^{\sqrt{z}} \frac{1}{\pi \varrho} e^{-\frac{r^2}{\varrho}} r dr d\theta = \frac{2\pi}{\pi \varrho} \int_0^{\sqrt{z}} r e^{-\frac{r^2}{\varrho}} dr. \quad (39)$$

By applying the substitution $u = r^2$ and $du = 2r dr$, the integral is simplified to

$$F_Z(z) = \frac{1}{\varrho} \int_0^z e^{-\frac{u}{\varrho}} du = 1 - e^{-\frac{z}{\varrho}}. \quad (40)$$

Taking the derivative of the CDF with respect to z , the PDF of the signal power is obtained as

$$f_Z(z) = \frac{d}{dz} F_Z(z) = \frac{1}{\varrho} e^{-\frac{z}{\varrho}}, \quad z \geq 0. \quad (41)$$

This precisely matches the PDF of an exponential distribution with a rate parameter of $\lambda = 1/\varrho$. Therefore, $|y_{n,i,t}|^2 \sim \text{Exp}(1/\varrho)$. This completes the proof of Lemma 2.

Proof of Theorem 1

The lemmas utilized in the following proof are derived from literature [32, Sec. E]. Hence, this section outlines only the primary steps. For the sake of simplicity, we consider the case where the number of scattering clusters is $L = 1$.

To establish the minimax optimality of the G-Exp-TS scheme for XL-MIMO systems, we analyze the cumulative regret per subarray. The total system regret is defined as $Q_{\text{total}}^{\text{reg}}(T) = \sum_{n=1}^N Q_n^{\text{reg}}(T)$.

We categorize the sub-optimal codewords into disjoint sets $F_{n,j} = \{i \in [M_n] \mid 2^{-(j+1)} \leq \delta_{n,i} < 2^{-j}\}$ based on their expected gain gaps $\delta_{n,i} = \mu_{n,1} - \mu_{n,i}$. To delineate the dominant regret boundaries, we introduce a critical separation threshold $\Delta_n^* = 4\sqrt{V_E M_n/T}$. The regret decomposition for subarray n can thus be reformulated as

$$\begin{aligned} Q_n^{\text{reg}}(T) &= \sum_{i: \delta_{n,i} > 0} \delta_{n,i} \cdot \mathbb{E}[C_{n,i,T}] \\ &\leq \sum_{i: \delta_{n,i} > \Delta_n^*} \delta_{n,i} \cdot \mathbb{E}[C_{n,i,T}] + \max_{i: \delta_{n,i} \leq \Delta_n^*} \delta_{n,i} \cdot T \\ &< \sum_{j < \hat{\omega}_n} \sum_{i \in F_{n,j}} 2^{-j} \cdot \mathbb{E}[C_{n,i,T}] + \Delta_n^* T, \end{aligned} \quad (42)$$

where the resolution depth is truncated at $\hat{\omega}_n = \frac{1}{2} \log_2(T/(V_E M_n)) - 3$. For each valid scale $j < \hat{\omega}_n$, we define the localized error margin as $\hat{\xi}_j = 2^{-j-2}$. The expected pulls $\mathbb{E}[C_{n,i,T}]$ are subsequently analyzed by decomposing the selection events into two primary risk types.

Expected Pulls under Sub-optimal Over-estimation (Type-I)

Type-I arises when a sub-optimal codeword $i \in F_{n,j}$ is selected because its index is over-estimated. For any $\hat{\zeta}_i > 0$ such that $\hat{\xi}_j + \hat{\zeta}_i < \delta_{n,i}$, and a constant $\kappa_{n,i}$ related to the noise variance, the expected pulls under event $E_{n,i,\hat{\xi}_j}^c(t)$ are bounded via the KL-divergence properties [32, Lemma E.3]

$$\begin{aligned} \text{Type-I} &= \sum_{i \in F_{n,j}} \mathbb{E} \left[\sum_t \mathbb{I}\{i \in \mathcal{A}_{n,t}, E_{n,i,\hat{\xi}_j}^c(t)\} \right] \\ &\stackrel{(a)}{\leq} \sum_{i \in F_{n,j}} \left(1 + \frac{2V_E}{\hat{\zeta}_i^2} + \frac{V_E}{\hat{\xi}_j^2} + \frac{\kappa_{n,i} \log(T\hat{\xi}_j^2/(V_E M_n))}{D_{KL}(\mu_{n,i} + \hat{\zeta}_i \|\mu_{n,1} - \hat{\xi}_j)} \right). \end{aligned} \quad (43)$$

In step (a) [32, Lemma E.3], the exponential decay property of KL divergence is utilized to restrict the over-sampling of sub-optimal arms, while the M_n term in the denominator precisely cancels the Union Bound penalty on the multi-armed state space.

Expected Pulls under Optimal Under-estimation (Type-II)

Type-II occurs when the optimal beam is under-estimated, i.e., $\beta_{n,1}(t) \leq \mu_{n,1} - \hat{\xi}_j$. We define $Q_{n,1,s}(\hat{\xi}_j)$ as the anti-concentration probability. Using the peeling technique over the interval $\hat{H}_s = (\mu_{n,1} - \hat{\xi}_j - \alpha_s, R_{max}]$, we bound the pulls under event $E_{n,i,\hat{\xi}_j}(t)$ as

$$\begin{aligned} \text{Type-II} &= \sum_{i \in F_{n,j}} \mathbb{E} \left[\sum_t \mathbb{I}\{i \in \mathcal{A}_{n,t}, E_{n,i,\hat{\xi}_j}(t)\} \right] \\ &\stackrel{(b)}{\leq} \sum_{s=1}^{\hat{J}_j} \mathbb{E} \left[\left(\frac{1}{Q_{n,1,s}(\hat{\xi}_j)} - 1 \right) \mathbb{I}\{\hat{\mu}_{n,1,s} \in \hat{H}_s\} \right] + \Theta \left(\frac{V_E M_n}{\hat{\xi}_j^2} \right). \end{aligned} \quad (44)$$

In step (b) [32, Lemma E.1], using the peeling technique, the cases where the empirical mean falls outside the high-probability confidence interval are isolated, and the anti-concentration probability $Q_{n,1,s}$ is utilized to control the dominant term.

Integrating the exploration probability $1/M_n$ of G-Exp-TS to offset the exponential anti-concentration term, we have

$$\text{Type-II} \stackrel{(c)}{=} O \left(\frac{V_E M_n \log(T\hat{\xi}_j^2/(M_n V_E))}{\hat{\xi}_j^2} \right). \quad (45)$$

Furthermore, (c) by leveraging the exploration probability $1/M_n$ of G-Exp-TS to offset the exponential term, the anti-concentration bound is integrated and scaled to eliminate the impact of constant terms [32, Lemma E.2].

Convergence and Final Bound

By substituting the Type-I and Type-II bounds back into the regret formula and setting $\hat{\zeta}_i = \hat{\xi}_j = \delta_{n,i}/4$, the summation over $j < \hat{\omega}_n$ forms a convergent series

$$\begin{aligned}
Q_n^{\text{reg}}(T) &\leq \sum_{j < \hat{\omega}_n} \sum_{i \in F_{n,j}} \hat{\xi}_j \cdot \mathbb{E}[C_{n,i,T}] + 4\sqrt{V_E M_n T} \\
&= O \left(\sum_{j < \hat{\omega}_n} \frac{M_n V_E \log(T \hat{\xi}_j^2 / (M_n V_E))}{\hat{\xi}_j} + \sqrt{V_E M_n T} + \sum_{i \geq 2} \delta_{n,i} \right) \\
&= O \left(8\sqrt{V_E M_n T} \cdot \sum_{k=0}^{\infty} \frac{\log 64 + k \log 2}{2^k} + \sum_{i \geq 2} \delta_{n,i} \right) \\
&= O \left(\sqrt{V_E M_n T} + \sum_{i \geq 2} \delta_{n,i} \right).
\end{aligned} \tag{46}$$

Finally, aggregating over N subarrays leads to the total regret

$$Q_{\text{total}}^{\text{reg}}(T) = \sum_{n=1}^N Q_n^{\text{reg}}(T) = O \left(N\sqrt{V_E M T} + \sum_{n,i \geq 2} \delta_{n,i} \right). \tag{47}$$

This completes the proof of the minimax optimality.

References

- [1] Tariq, F., Khandaker, M.R.A., Wong, K.-K., Imran, M.A., Bennis, M., Debbah, M.: A speculative study on 6g. *IEEE Wireless Communications* **27**(4), 118–125 (2020) <https://doi.org/10.1109/MWC.001.1900488>
- [2] Zeng, Y., Chen, J., Xu, J., Wu, D., Xu, X., Jin, S., Gao, X., Gesbert, D., Cui, S., Zhang, R.: A tutorial on environment-aware communications via channel knowledge map for 6g. *IEEE Communications Surveys & Tutorials* **26**(3), 1478–1519 (2024) <https://doi.org/10.1109/COMST.2024.3364508>
- [3] Parra-Ullauri, J.M., Zhang, X., Bravalheri, A., Moazzeni, S., Wu, Y., Nejabati, R., Simeonidou, D.: Federated analytics for 6g networks: Applications, challenges, and opportunities. *IEEE Network* **38**(2), 9–17 (2024) <https://doi.org/10.1109/MNET.2024.3355218>
- [4] Na, M., Lee, J., Choi, G., Yu, T., Choi, J., Lee, J., Bahk, S.: Operator’s perspective on 6g: 6g services, vision, and spectrum. *IEEE Communications Magazine* **62**(8), 178–184 (2024) <https://doi.org/10.1109/MCOM.001.2400060>
- [5] González-Prelcic, N., Furkan Keskin, M., Kaltiokallio, O., Valkama, M., Dardari, D., Shen, X., Shen, Y., Bayraktar, M., Wymeersch, H.: The integrated sensing

- and communication revolution for 6g: Vision, techniques, and applications. *Proceedings of the IEEE* **112**(7), 676–723 (2024) <https://doi.org/10.1109/JPROC.2024.3397609>
- [6] Psomas, C., Ntougias, K., Shanin, N., Xu, D., Mayer, K., Tran, N.M., Cottatellucci, L., Choi, K.W., Kim, D.I., Schober, R., Krikidis, I.: Wireless information and energy transfer in the era of 6g communications. *Proceedings of the IEEE* **112**(7), 764–804 (2024) <https://doi.org/10.1109/JPROC.2024.3395178>
 - [7] Gao, T., Song, Y., Liu, C., Yin, Z., Cheng, N., Ge, D.: Hybrid-field-aware two-stage beamforming scheme for xl-mimo systems. *IEEE Internet of Things Journal* **12**(13), 24119–24130 (2025) <https://doi.org/10.1109/JIOT.2025.3555635>
 - [8] Luo, J., Fan, J., Xie, K., Shi, X.: Efficient hybrid near- and far-field beam training for xl-mimo communications. *IEEE Transactions on Vehicular Technology* **73**(12), 19785–19790 (2024) <https://doi.org/10.1109/TVT.2024.3447129>
 - [9] Chen, K., Qi, C., Dobre, O.A., Ye Li, G.: Triple-refined hybrid-field beam training for mmwave extremely large-scale mimo. *IEEE Transactions on Wireless Communications* **23**(8), 8556–8570 (2024) <https://doi.org/10.1109/TWC.2024.3351712>
 - [10] Gao, Z., Zhou, X., Ning, B., Su, Y., Qin, T., Niyato, D.: Integrated location sensing and communication for ultra-massive mimo with hybrid-field beam-squint effect. *IEEE Journal on Selected Areas in Communications* **43**(4), 1387–1404 (2025) <https://doi.org/10.1109/JSAC.2025.3531551>
 - [11] Lin, Z., Wang, X., Zeng, D., Wu, Y.: Deep learning assisted hybrid beamforming for thz hybrid near and far field communications. In: 2025 IEEE 25th International Conference on Communication Technology (ICCT), pp. 752–756 (2025). <https://doi.org/10.1109/ICCT67417.2025.11374102>
 - [12] Weng, C., Guo, X., Guo, Y., Wang, Y.: Wavenumber domain beam training in xl-mimo systems: Unifying far-field and near-field. *IEEE Transactions on Wireless Communications* **25**, 2183–2196 (2026) <https://doi.org/10.1109/TWC.2025.3595204>
 - [13] Kurt, A., Guvensen, G.M.: Robust statistical beamforming with multi-cluster tracking for time-varying massive mimo. *IEEE Transactions on Vehicular Technology* **73**(3), 3499–3515 (2024) <https://doi.org/10.1109/TVT.2023.3323792>
 - [14] Jiang, P., Fu, J., Wang, Y., Zhu, P.: Statistical channel state information aided local precoding optimization for cell free massive mimo systems. In: ICC 2025 - IEEE International Conference on Communications, pp. 4842–4847 (2025). <https://doi.org/10.1109/ICC52391.2025.11161403>

- [15] Shi, X., Wang, J., Sun, Z., Song, J.: Spatial-chirp codebook-based hierarchical beam training for extremely large-scale massive mimo. *IEEE Transactions on Wireless Communications* **23**(4), 2824–2838 (2024) <https://doi.org/10.1109/TWC.2023.3303229>
- [16] Weng, C., Guo, X., Wang, Y.: Near-field beam training with hierarchical codebook: Two-stage learning-based approach. *IEEE Transactions on Vehicular Technology* **73**(9), 14003–14008 (2024) <https://doi.org/10.1109/TVT.2024.3391878>
- [17] Jiang, G., Qi, C.: Near-field beam training based on deep learning for extremely large-scale mimo. *IEEE Communications Letters* **27**(8), 2063–2067 (2023) <https://doi.org/10.1109/LCOMM.2023.3289513>
- [18] Liu, Y., Li, M., Deng, W., Zhao, M.-J.: Adaptive position-aware near-field beam training for millimeter-wave xl-mimo. *IEEE Internet of Things Journal* **12**(18), 38185–38197 (2025) <https://doi.org/10.1109/JIOT.2025.3586210>
- [19] Hojatian, H., Nadal, J., Frigon, J., Leduc-Primeau, F.: Flexible unsupervised learning for massive mimo subarray hybrid beamforming. In: *GLOBECOM 2022 - 2022 IEEE Global Communications Conference*, pp. 3833–3838 (2022). <https://doi.org/10.1109/GLOBECOM48099.2022.10001155>
- [20] Chen, K., Qi, C., Wang, C.-X., Li, G.Y.: Beam training and tracking for extremely large-scale mimo communications. *IEEE Transactions on Wireless Communications* **23**(5), 5048–5062 (2024) <https://doi.org/10.1109/TWC.2023.3324176>
- [21] Yu, J., Lee, C.: Beam training for partially-connected structure in xl-mimo systems with spatial non-stationarity. In: *2025 1st International Conference on Consumer Technology (ICCT-Pacific)*, pp. 1–2 (2025). <https://doi.org/10.1109/ICCT-Pacific63901.2025.11012794>
- [22] Takigawa, M., Kataoka, R., Kanno, I., Kishi, Y.: Beam training of los-mimo systems using subarray-based beamforming in the presence of ground reflection. In: *2024 18th European Conference on Antennas and Propagation (EuCAP)*, pp. 1–5 (2024). <https://doi.org/10.23919/EuCAP60739.2024.10501374>
- [23] Zhou, C., Wu, C., You, C., Zhou, J., Shi, S.: Near-field beam training with sparse dft codebook. *IEEE Transactions on Communications* **73**(6), 4394–4408 (2025) <https://doi.org/10.1109/TCOMM.2024.3511702>
- [24] Jiang, Y., Liu, C., Song, Y., Liang, H.: Near-field beam training for xl-mimo systems: A multi-armed bandit-based approach. *IEEE Transactions on Vehicular Technology*, 1–6 (2026) <https://doi.org/10.1109/TVT.2026.3656274>
- [25] Dinh-van, S., Mo, Y.K., Zidan, J., Fowler, D.S., Evans, A., Higgins, M.D., Trinh,

- V.C.: Rapid beam training at terahertz frequency with contextual multi-armed bandit learning. In: 2024 IEEE International Conference on Industrial Technology (ICIT), pp. 1–7 (2024). <https://doi.org/10.1109/ICIT58233.2024.10540692>
- [26] Im, C., Lee, C.: Low-complexity hybrid beamforming for multiuser millimeter wave systems with collaborative learning. *IEEE Transactions on Wireless Communications* **23**(9), 12115–12125 (2024) <https://doi.org/10.1109/TWC.2024.3387988>
- [27] Liang, H., Liu, C., Song, Y., Yin, Z., Wang, G.: Joint two stage beamforming and user scheduling for leo satellite communications via hierarchical multi-armed bandit. *IEEE Transactions on Cognitive Communications and Networking* **12**, 1918–1932 (2026) <https://doi.org/10.1109/TCCN.2025.3602862>
- [28] Arai, M., Yamagami, T., Mihana, T., Horisaki, R., Hasegawa, M.: Two-dimensional beam selection by multiarmed bandit algorithm based on a quantum walk. *IEEE Transactions on Quantum Engineering* **6**, 1–16 (2025) <https://doi.org/10.1109/TQE.2025.3555145>
- [29] Dong, Z., Li, X., Zeng, Y., Jin, S., Jiang, T.: Near-field spatial correlation for multi-path xl-array communications with partial visibility. In: GLOBECOM 2023 - 2023 IEEE Global Communications Conference, pp. 1525–1530 (2023). <https://doi.org/10.1109/GLOBECOM54140.2023.10437851>
- [30] Farayola, P.O., Chaganti, S.K., Obaidi, A.O., Sheikh, A., Ravi, S., Chen, D.: Quantile – quantile fitting approach to detect site to site variations in massive multi-site testing. In: 2020 IEEE 38th VLSI Test Symposium (VTS), pp. 1–6 (2020). <https://doi.org/10.1109/VTS48691.2020.9107616>
- [31] Song, Y., Liu, C., Li, H., Yin, Z., Liang, H., Wang, H.: Spectral-efficient tsb scheme for multiple scatterer clusters channel in massive mimo. *IEEE Communications Letters* **27**(11), 3058–3062 (2023) <https://doi.org/10.1109/LCOMM.2023.3315855>
- [32] Jin, T., Xu, P., Xiao, X., Anandkumar, A.: Finite-time regret of thompson sampling algorithms for exponential family multi-armed bandits. *Advances in Neural Information Processing Systems* **35**, 38475–38487 (2022)