Grid-Assisted Kalman Filter-based Beam Tracking for Dynamic Near-Field XL-MIMO

Jikun Zhu[1], Zheng Wang[1], Zhen Gao[2], Yongming Huang[1]
[1]*School of Information Science and Engineering, Southeast University, Nanjing, China*
[2]*School of Information and Electronics, Beijing Institute of Technology, Beijing, China*
Email: jikun_zhu@seu.edu.cn, wznuaa@gmail.com, gaozhen16@bit.edu.cn, huangym@seu.edu.cn

***Abstract*—Dynamic near-field beam training in extremely large-scale MIMO (XL-MIMO) systems is challenging due to the high overhead of jointly searching angle and distance domains. This paper proposes a novel grid-assisted Kalman filter-based beam tracking (GKBT) scheme for efficient beam alignment in dynamic near-field XL-MIMO. GKBT recursively tracks the time-varying channel parameters as the user moves by combining Kalman filter prediction with a low-overhead local grid maximum-likelihood (ML) search. At each time slot, the Kalman filter predicts the beam direction. Then a small local codebook search around the prediction yields an ML estimate to update the filter state. By confining training to a local beam subset, GKBT significantly reduces per-slot overhead compared to the repeated global beam training. Simulation results demonstrate that GKBT consistently achieves near-optimal beamforming performance with only a fraction of the training overhead required by conventional retraining schemes.**

***Index Terms*—Extremely large-scale MIMO (XL-MIMO), near field communications, beam tracking, Kalman filtering, local grid maximum-likelihood (ML).**

I. Introduction

With the evolution toward 6G, XL-MIMO has emerged as a key technology for future wireless systems due to its high spatial resolution and spectral efficiency [1]. However, when the array aperture grows large, the transition from far-field planar wave propagation to near-field spherical wave propagation renders traditional far-field beamforming and training methods ineffective [2]. Consequently, near-field beamforming and precoding techniques have been actively studied to achieve precise beam alignment in XL-MIMO systems.

Early research in [3] proposes polar-domain codebooks employing uniform angle sampling and non-uniform distance sampling. While these codebooks can accurately represent near-field steering vectors, their high dimensionality and training overhead make exhaustive search impractical. To further reduce training overhead, a fast near-field training scheme in [4] first uses an angle-only far-field codebook to identify candidate angles and then estimates distance, yet its achievable-rate performance under complicated channel conditions often falls short of expectations. Moreover, the time-delay beamforming method in [5] exploits the frequency-dependent beam split in wideband systems to cover different distances at different frequencies, achieving extremely low training overhead and high performance. However, such methods require dedicated hardware and are costly and power-consuming. Meanwhile, a novel near-field 2D hierarchical beam training approach in [6] strikes a balance between high achievable rate and moderate overhead, but its iterative codeword design is computationally expensive. To reduce complexity, a cluster-based hierarchical beam training (CHB) scheme is proposed in [7]. Despite its advantages in both overhead and achievable rate, CHB still remains far from optimal performance and is primarily designed for static channel conditions. However, in dynamic 6G user scenarios, the near-field channel parameters vary over time due to user mobility. This necessitates continuous beam tracking instead of repeatedly re-training.

To further improve beam training performance and to accommodate dynamic near-field channels, we propose a grid-assisted Kalman filter (KF)-based beam tracking (GKBT) scheme for dynamic near-field XL-MIMO under a digital precoding architecture. In GKBT, a Kalman filter is employed to recursively track the time-varying near-field parameters, while a local-grid maximum-likelihood (ML) search procedure is used to obtain reliable measurements for filter correction. Simulation results demonstrate that GKBT offers a significant performance advantage over other low-overhead beam training algorithms. By exploiting the predictability of dynamic channels, GKBT consistently tracks near-optimal beams with only a few local observations. This achieves a better trade-off between performance and training overhead.

II. System Model

As illustrated in Fig. 1, we consider the downlink beam training of a narrow-band XL-MIMO communication system. In this setup, base station (BS) is equipped with a uniform linear array (ULA) comprising N antennas for communication with a single-antenna use equipment (UE).

When the distance between BS and UE is smaller than the Rayleigh distance $Z = \frac{2D^2}{\lambda}$ with D and λ denoting the antenna array aperture and the carrier wavelength respectively, UE operates in the near-field region [7]. Consequently, the traditional far-field plane-wave assumption no longer holds, and the channel is characterized by a spherical-wave model:

$$\mathbf{h} = \sqrt{N} g_l \mathbf{b}\left(\theta, r\right), \tag{1}$$

where g_l is the complex-valued channel gain. Here, $\mathbf{b}\left(\theta, r\right)$ represents the near-field steering vector:

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

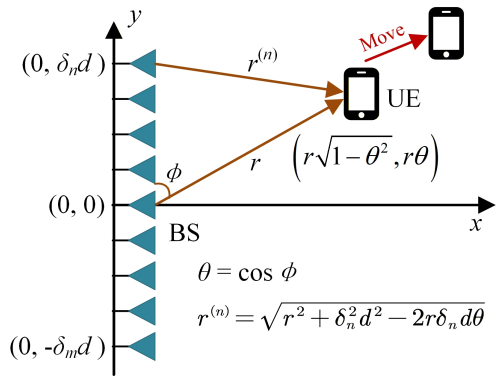

Fig. 1. A downlink near-field XL-MIMO communication system.

where θ indicates the spatial angle at BS, and r represents the distance from UE to the center of the antenna array. The distance between the n-th antenna at BS and UE is given by $r^{(n)} = \sqrt{r^2 + \delta_n^2 d^2 - 2r\delta_n d\theta}$, where $d = \frac{\lambda}{2}$ is the antenna spacing and $\delta_n = \frac{2n-N-1}{2}$ with $n = 0, 1, 2, ..., N$. Under the downlink near-field channel model, the received signal at UE is expressed as:

$$y = \mathbf{h}^H \mathbf{v} s + n_0 = \sqrt{N} g_l \mathbf{b}^H (\theta, r) \mathbf{v} s + n_0, \tag{3}$$

where $\mathbf{v} \in \mathbb{C}^{N\times 1}$ indicates the beamforming vector at BS, $s \in \mathbb{C}$ represents the transmitted symbol and n_0 is the received additive white Gaussian noise (AWGN) component with power σ^2. As such, given perfect CSI information (θ, r), the optimal XL-MIMO BS beamforming vector can be obtained as $\mathbf{v}^* = \mathbf{b}(\theta, r)$ [3].

In dynamic user scenarios, the motion of UE causes near-field parameters vary over time. We consider T consecutive time frames and define the historical channel sequence $\mathcal{H}$ as:

$$\mathcal{H} \triangleq \{\mathbf{h}_t = \mathbf{h}(\theta_t, r_t) | t = 1, ..., T\}, \tag{4}$$

where t denotes the time-frame index, $\mathbf{h}_t$ is the channel vector at frame t, and (θ_t, r_t) are the corresponding angular and distance parameters. At the prediction frame $T+1$, the received signal can be expressed based on (3) as:

$$y_{T+1} = \mathbf{h}_{T+1}^H \mathbf{v} s + n_0. \tag{5}$$

Assume that beamforming is performed using a predefined codebook. The design principle of the codebook is to cover the largest possible spatial range, enabling precise beam alignment in different directions. The current effective approach is to construct the polar-domain codebook $\mathbf{M} \in \mathbb{C}^{N\times(NS_N)}$ by including a large number of the optimal beamforming vectors from different sampling points (θ_n, r_n^s):

$$\mathbf{M} = \left[\mathbf{b}(\theta_1, r_1^1), ..., \mathbf{b}(\theta_N, r_N^1), ..., \mathbf{b}\left(\theta_N, r_N^{S_N}\right)\right], \tag{6}$$

where N represents the number of antennas, and S_N denotes the number of distance rings in polar sampling. The angle θ is uniformly sampled as $\theta_n = \frac{2n-N+1}{N}, n = 0, 1, ..., N$, while the distance r is non-uniformly sampled with a threshold distance Z_Δ [3] according to formula $r_n^s = \frac{1}{s} Z_\Delta \left(1 - \theta_n^2\right), s = 1, 2, ..., S_N$. Here, n indicates the different angle directions used for sampling, while s indicates the various rings used for distance sampling. To be specific, each column of $\mathbf{M}$ represents a codeword $\mathbf{w}_{s,n} = \mathbf{b}(\theta_n, r_n^s)$. The beam selection problem can be formulated as finding the beamforming vector that maximizes achievable rate:

$$\max_{\mathbf{v}} E[R(\mathbf{v})] = \max_{\mathbf{v}} \left[\log_2\left(1 + \rho |\mathbf{h}_{T+1}^H \mathbf{v}|^2\right)\right], \tag{7}$$

where ρ denotes the transmit signal-to-noise ratio (SNR) coefficient. However, since the polar-domain codebook contains $N \times S_N$ codewords, repeatedly performing exhaustive search for every frame in dynamic XL-MIMO scenarios would incur prohibitive overhead. This challenge motivates the development of low-overhead beam tracking designs.

III. Proposed Grid-Assisted Kalman Filtering Based Near-Field Beam Tracking

This section starts by reviewing a previously proposed low-overhead near-field beam training algorithm CHB. Building on the dynamic channel scenario described in the preceding sections, we then develop the proposed GKBT scheme. Finally, we compare the training overhead of GKBT with other representative beam training approaches.

A. Cluster-based Hierarchical Beam Training

To reduce computational complexity and training overhead of near-field beam training in XL-MIMO, CHB leverages clustering to construct a low-dimensional polar-domain codebook and further adopts hierarchical beam training to decrease beam search region [7]. Starting from the polar-domain near-field codebook $\mathbf{M}$, CHB applies k-means clustering separately on the sampled angles and ranges to obtain k_θ angular cluster centers μ_j^θ and k_r distance cluster centers μ_j^r into j class. By pairing these cluster center as $\left(\mu_m^\theta, \mu_n^r\right)$, the near-field steering vector $\mathbf{b}\left(\mu_m^\theta, \mu_n^r\right)$ is computed via (2) to serve as the near-field codeword $\mathbf{w}_{mn}$. Finally, these codewords are combined to form a cluster-based hierarchical near-field codebook $\mathbf{W}$ of size $k_\theta \times k_r$:

$$\mathbf{W} = \left\{\mathbf{w}_{m,n} = \mathbf{b}\left(\mu_m^\theta, \mu_n^r\right) | m = 1, ..., k_\theta, n = 1, ..., k_r\right\}. \tag{8}$$

After the design of codebook, CHB organizes $\mathbf{W}$ into multiple layers with increasing resolution. At a given layer, the BS probes candidate codewords and selects the one maximizing the received power, which can be expressed as:

$$(m^\star, n^\star) = \arg\max_{m,n} \left|\mathbf{h}^H \mathbf{w}_{m,n}\right|^2. \tag{9}$$

The selected center pair $(\mu_{m^\star}^\theta, \mu_{n^\star}^r)$ determines a refined polar-range region for the next layer searches. Such hierarchical search avoids exhaustive scanning over the full polar grid and substantially reduces training overhead.

However, in dynamic channels, CHB is unable to capture the motion state of UE. Consequently, the resulting beam training performance is essentially the same as that in a randomly generated static channel. Moreover, CHB must re-initiate the training and calibration procedure from scratch at every frame during the UE's motion, which leads to substantial overall overhead and yields unsatisfactory final performance.

B. Kalman Filter Beam Tracking with Local ML Search

We first apply CHB to perform a coarse beam alignment, which serves as the starting point for the subsequent study. In dynamic near-field XL-MIMO scenarios, the near-field parameters (θ, r) evolve over time due to user mobility. Rather than treating a dynamic link as a collection of independent random static snapshots, we explicitly exploit the temporal predictability of (θ, r) and perform low-overhead tracking via Kalman filtering assisted by a local-grid ML observation.

a) Dynamic State and Observation Model: Let the discrete time index be k with slot duration ΔT. We define the user motion state $\mathbf{x}_k$ as a constant-velocity model in polar domain:

$$\mathbf{x}_k = \left[r_k, \dot{r}_k, \theta_k, \dot{\theta}_k\right]^T, \tag{10}$$

where r_k and θ_k denote the distance and angle at time k with their corresponding temporal variation rates $\dot{r}_k$ and $\dot{\theta}_k$. Under the widely used constant-velocity assumption, the state transition is

$$\mathbf{x}_k = \mathbf{F}\mathbf{x}_{k-1} + \mathbf{q}_{k-1}, \tag{11}$$

with

$$\mathbf{F} = \begin{bmatrix} 1 & \Delta T & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & \Delta T \\ 0 & 0 & 0 & 1 \end{bmatrix}, \tag{12}$$

where $\mathbf{q}_{k-1} \sim \mathcal{N}(\mathbf{0}, \mathbf{Q})$ and $\mathbf{Q} \in \mathbb{R}^{4\times 4}$ are the process-noise covariance accounting for model mismatch in the assumed motion dynamics as well as random perturbations in the target motion. At each time k, BS obtains a short observation $(\tilde{\theta_k}, \tilde{r_k})$ and feeds it into the KF as a measurement:

$$\mathbf{z}_k = \begin{bmatrix} \tilde{r}_k \\ \tilde{\theta}_k \end{bmatrix} = \mathbf{A}\mathbf{x}_k + \mathbf{u}_k, \tag{13}$$

where $\mathbf{u}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{R})$ denotes measurement noise with the measurement noise covariance matrix $\mathbf{R} \in \mathbb{R}^{2\times 2}$ and

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}. \tag{14}$$

Under the digital precoding with a single-antenna UE considered in this paper, only a single radio frequency (RF) chain is available at UE side, which precludes multi-dimensional observations at BS. Therefore, the key challenge is to construct the measurement $(\tilde{\theta_k}, \tilde{r_k})$ with small overhead [8], which we address via a local-grid ML search in the next paragraph.

b) Local-Grid Assisted ML Measurement: We leverage the KF prediction to confine the search space to a small neighborhood and construct a low-overhead local-grid ML measurement. Specifically, centered at prior prediction $(\hat{\theta}_{k-1}, \hat{r}_{k-1})$ of the near-field parameters at slot k, we define a 2-dimensional (2D) local candidate set:

$$\begin{aligned} \mathcal{G}_k \triangleq \Big\{ (\theta, r) \,\Big|\, & \theta = \hat{\theta}_{k-1} + i\Delta_\theta,\ r = \hat{r}_{k-1} + j\Delta_r \Big\} \\ & \text{with } i \in \{-I, \ldots, I\},\ j \in \{-J, \ldots, J\}, \end{aligned} \tag{15}$$

where Δ_θ and Δ_r are the grid resolutions, I and J control the polar search span. The number of probed candidates per slot is $|\mathcal{G}_k| = (2I+1)(2J+1)$, which directly determines the tracking overhead. For each candidate $(\theta, r) \in \mathcal{G}_k$, BS forms the corresponding near-field beam:

$$\mathbf{v}(\theta, r) = \frac{\mathbf{b}(\theta, r)}{\| \mathbf{b}(\theta, r) \|}, \tag{16}$$

and evaluates the received pilot metric. The ML estimate is obtained by maximizing the beamformed received power:

$$(\tilde{\theta_k}, \tilde{r_k}) = \arg \max_{(\theta, r) \in \mathcal{G}_k} \left| \mathbf{h}_k^H \mathbf{v}(\theta, r) \right|^2 \tag{17}$$

where $\mathbf{h}_k$ denotes the near-field channel model at time k. Finally, the local ML output serves as the KF measurement $\mathbf{z}_k$ will be fed into the KF correction step to refine $\hat{\mathbf{x}}_{k|k}$ and update the beamforming vector for data transmission. In this way, KF prediction reduces the search region, while the local ML measurement anchors the state estimate using only a small number of pilots.

c) Kalman Recursion and Beam Tracking: Based on the state transition and observation models, the Kalman filter recursively estimates the user state. At each time instant k, the KF proceeds in two steps prediction and correction. Let $\hat{\mathbf{x}}_{k|k-1}$ and $\hat{\mathbf{x}}_{k|k}$ denote the predicted and the filtered state estimates at slot k respectively, the prediction state can be expressed based on velocity transition matrix $\mathbf{F}$ in (12) as:

$$\hat{\mathbf{x}}_{k|k-1} = \mathbf{F}\hat{\mathbf{x}}_{k-1|k-1}. \tag{18}$$

Meanwhile, the state estimation error covariance matrix $\mathbf{P}$ of KF is updated accordingly:

$$\mathbf{P}_{k|k-1} = \mathbf{F}\mathbf{P}_{k-1|k-1}\mathbf{F}^T + \mathbf{Q}. \tag{19}$$

Upon obtaining the measurement $\mathbf{z}_k$, the correction step is performed using the linear observation matrix $\mathbf{A}$, where the measurement residual is given by:

$$\mathbf{y}_k = \mathbf{z}_k - \mathbf{A}\hat{\mathbf{x}}_{k|k-1}, \tag{20}$$

and the residual covariance is

$$\mathbf{S}_k = \mathbf{A}\mathbf{P}_{k|k-1}\mathbf{A}^T + \mathbf{R}. \tag{21}$$

Algorithm 1 Grid-Assisted KF-based Beam Tracking

Input: N, CHB-generated codebook $\mathbf{W}$, ΔT, $\mathbf{Q}$, $\mathbf{R}$, local-grid parameter $(\Delta_\theta, \Delta_r, I, J)$, tracking window length K

1: Initialization: Perform CHB at $k = 1$ over the codebook $\mathbf{W}$ to obtain $\mathbf{v}_1$ and the initial position estimate $\hat{\mathbf{x}}_{1|1}$. Let the state estimation error covariance matrix as $\mathbf{P}_{1|1} = \mathbf{P}_0$
2: **for** $k = 2, ..., K$ **do**
3: calculate the priori estimate $\hat{\mathbf{x}}_{k|k-1}$ by (18) and (19)
4: construct the candidate set $\mathcal{G}_k$ by (15)
5: obtain $(\tilde{\theta_k}, \tilde{r_k})$ by (17) and $\mathbf{z}_k$ by (13)
6: calculate residual $\mathbf{y}_k$ and $\mathbf{S}_k$ by (20) and (21)
7: calculate Kalman gain $\mathbf{K}_k$ by (22)
8: calculate state estimate $\hat{\mathbf{x}}_{k|k}$ and $\left(\hat{\theta}_k, \hat{r}_k\right)$ by (23)
9: update the beamforming vector $\mathbf{v}_k$ by (16)
10: **end for**

Output: Tracked state $\hat{\mathbf{x}}_{k|k}$ and beamforming vector $\mathbf{v}_k$

TABLE I
NEAR-FIELD BEAM TRAINING METHODS AND OVERHEAD IN DIFFERENT SCENARIOS

Methods	Training Overhead	Scenario I	Scenario II
Near-field Exhaustive [3]	$K\mathcal{O}(NS_N)$	$2048K$	$8192K$
Near-field Hierarchical [6], [7]	$K\mathcal{O}\left(k_\theta^{(1)}k_r^{(1)}\right)$	$64K$	$192K$
CHB-initialized GKBT	$\mathcal{O}\left(k_\theta^{(1)}k_r^{(1)}\right)+(K-1)\mathcal{O}\left((2I+1)(2J+1)\right)$	$64+9(K-1)$	$192+25(K-1)$

Scenario I: $N=256$, $S_N=8$, $k_\theta^{(1)}=16$, $k_r^{(1)}=4$, $I=J=1$; **Scenario II:** $N=512$, $S_N=16$, $k_\theta^{(1)}=24$, $k_r^{(1)}=8$, $I=J=2$.

The Kalman gain $\mathbf{K}_k \in \mathbb{R}^{4\times 2}$ can be then obtained as:

$$\mathbf{K}_k = \mathbf{P}_{k|k-1}\mathbf{A}^T\mathbf{S}_k^{-1} = \mathbf{P}_{k|k-1}\mathbf{A}^T\left(\mathbf{A}\mathbf{P}_{k|k-1}\mathbf{A}^T+\mathbf{R}\right)^{-1}. \tag{22}$$

Finally, the filtered estimate and its covariance are updated as:

$$\begin{aligned}\hat{\mathbf{x}}_{k|k} &= \hat{\mathbf{x}}_{k|k-1}+\mathbf{K}_k\left(\mathbf{z}_k-\mathbf{A}\hat{\mathbf{x}}_{k|k-1}\right),\\ \mathbf{P}_{k|k} &= (\mathbf{I}-\mathbf{K}_k\mathbf{A})\mathbf{P}_{k|k-1},\end{aligned} \tag{23}$$

continuously providing updated estimates of UE's current distance and angle. Intuitively, $\mathbf{K}_k$ balances the trust between the model prediction and the local ML measurement. When $\mathbf{R}$ is small, the filter relies more on $\mathbf{z}_k$; when $\mathbf{Q}$ is small, the filter relies more on the prediction. Using the estimated $(\hat{\theta}_k, \hat{r}_k)$ in $\hat{\mathbf{x}}_{k|k}$, we compute the beamforming vector $\mathbf{v}_k$ at the current time instant according to (16). Algorithm 1 summarizes the GKBT procedure, which starts from initialization and repeatedly updates the user location via Kalman filtering, enabling beam alignment under dynamic channels.

C. Overhead Analysis and Comparison

In randomly static near-field channels, each beam training round of a codebook-based scheme is typically dictated by the size of thecodebook. The exhaustive scheme needs to traverse the entire polar-domain codebook, incurring overhead of $\mathcal{O}(NS_N)$ [3]. From the hierarchical beam training perspective, the overhead is dictated by the initial low-resolution search. Take CHB as an example, its total training overhead is $\mathcal{O}\left(\sum\left(k_\theta^{(l)}k_r^{(l)}\right)\right)$, where $k_\theta^{(l)}k_r^{(l)}$ denotes the number of codewords to be searched at layer l. In many cases, this overhead can be approximated by only the first layer $\mathcal{O}\left(k_\theta^{(1)}k_r^{(1)}\right)$ [7]. Since the low-resolution codebook has a smaller dimension, the hierarchical approach significantly reduces training overhead compared with exhaustive search.

Unlike codebook-based beam training schemes, the overhead of GKBT mainly stems from a one-time beam alignment during initialization and the subsequent beam tracking cost. In dynamic near-field channels, we consider a tracking interval consisting of K consecutive time slots. GKBT incurs a one-time initialization overhead to obtain an initial alignment $\hat{\mathbf{x}}_{1|1}$, followed by a per-slot tracking overhead determined by the local-grid size. Specifically, the total overhead can be summarized as $\mathcal{O}\left(k_\theta^{(1)}k_r^{(1)}\right)+(K-1)\,\mathcal{O}\left((2I+1)(2J+1)\right)$. Meanwhile, codebook-based schemes must re-run beam training whenever the channel varies from slot to slot. Therefore, under the dynamic near-field channel model, the overhead of exhaustive search is $K\mathcal{O}(NS_N)$ and that of hierarchical training methods such as CHB is $K\mathcal{O}\left(k_\theta^{(1)}k_r^{(1)}\right)$. Such repeated re-training leads to rapidly increasing overhead over time. Moreover, since $(2I+1)(2J+1) \ll k_\theta^{(1)}k_r^{(1)} \ll NS_N$, GKBT achieves a significant overhead advantage over codebook-based schemes in dynamic scenarios. Table I summarizes the overhead comparison among these near-field beam training methods in different scenarios and the parameters are listed below the table. Although GKBT incurs additional overhead on top of the initialization stage in randomly static channels, it can exploit the temporal correlation of prior channel states to achieve better performance than conventional codebook-based beam training. Since the extra cost introduced by the local-grid search is marginal, the proposed GKBT strikes a favorable performance-overhead tradeoff in static near-field channels, while offering a pronounced advantage over existing methods in dynamic scenarios.

IV. SIMULATION RESULTS

In this section, we present simulation results to evaluate the effectiveness of the proposed GKBT scheme. We consider a dynamic narrowband XL-MIMO system with 512 antennas at BS. The system bandwidth is 100 MHz, and the carrier frequency is 50 GHz, which corresponds to a wavelength of 0.006 meters. Since the antenna spacing is half of the wavelength, the array aperture is $D = N\times\frac{\lambda}{2} = 1.536m$. The calculated Rayleigh distance is $Z = \frac{2D^2}{\lambda} = 786.432m$. In our simulations, the user distance ranges from $(20, 100)$, which is well within the Rayleigh distance, ensuring that all users operate in the near-field regime. Additionally, the angle range is $\theta \sim \mathcal{U}(-1,1)$, and the complex path gain g_l satisfies $g_l \sim \mathcal{CN}(0,1)$. Instead of relying on real measurement data, this work employs a physically consistent near-field channel model to generate large-scale samples through offline simulation, ensuring that the network can effectively cover the two-dimensional angle-distance search space within the entire Fresnel region. The mobility follows a constant-velocity model: a radial velocity $v_r \sim \mathcal{N}(0,\sigma_{v_r}^2)$ and an angular velocity $v_\theta \sim \mathcal{N}(0,\sigma_{v_\theta}^2)$, where $\sigma_{v_r}^2 = 0.4m/\text{slot}$ and $\sigma_{v_\theta} = 1^\circ$ per slot.

Fig. 2 shows the achievable rate versus system SNR for various schemes in XL-MIMO systems. As a performance upper bound, the near-field exhaustive search consistently achieves the

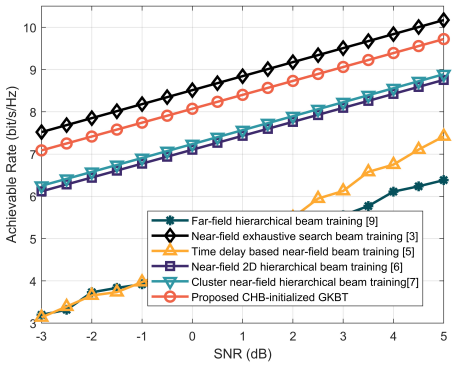

Fig. 2. Achievable sum-rate performance comparison versus SNR.

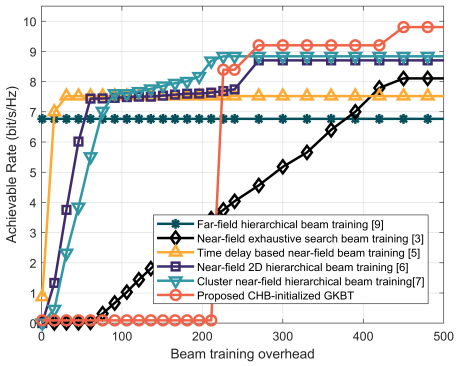

Fig. 3. Achievable sum-rate performance comparison versus overhead.

highest rate across the entire SNR range. The proposed CHB-initialized GKBT stays close to this upper bound and exhibits a stable, near-parallel SNR scaling trend. Importantly, the gap between GKBT and exhaustive search remains relatively small and does not amplify with SNR, which suggests that the proposed tracking mechanism is not dominated by noise-limited coarse alignment. Instead, it is primarily limited by the intentionally small local probing budget and the local-grid resolution, both of which are designed to control overhead. In contrast, the near-field 2D hierarchical [6] and CHB [7] methods also improve with SNR but consistently perform worse than GKBT. The fixed codebook structure limits their refinement capability unless additional layers or denser codewords are used. The far-field hierarchical beam training [8] performs substantially worse than all near-field methods and shows a weaker effective SNR gain. This is mainly due to the model mismatch between the far-field planar-wave steering assumption and the near-field spherical-wave channel [10].

Fig. 3 illustrates the achievable rate versus beam training overhead in random static near-field channels. The near-field 2D hierarchical and CHB approaches offer a reasonable trade-off by using multi-resolution codebooks, yet their performance is bounded by fixed layer resolutions. It should be noted that the overhead of the proposed GKBT consists of two distinct parts: a one-time CHB-based initial beam acquisition and the subsequent low-cost tracking updates. Therefore, the turning point of GKBT mainly corresponds to the completion of the initial alignment stage, rather than its steady-state tracking cost. From this perspective, Fig. 3 presents a conservative view of GKBT, because it includes the entry cost of beam acquisition in a static one-shot comparison. This enables it to reach the high-rate region with significantly reduced additional overhead.

V. Conclusion

In this paper, we propose a grid-assisted Kalman filter based beam tracking scheme for dynamic near-field XL-MIMO systems. GKBT performs cluster-based hierarchical beam training once for initialization, and then recursively tracks the time-varying near-field parameters by combining Kalman prediction with a low-overhead local-grid maximum-likelihood search as reliable observations. By confining per-slot probing to a small neighborhood around the predicted direction, GKBT substantially reduces the beam training overhead compared with repeated re-training based schemes, while maintaining accurate beam alignment. Simulation results verify that GKBT achieves near-optimal achievable rates with only a fraction of the training cost, demonstrating its effectiveness and practicality for real-time near-field beam alignment in dynamic XL-MIMO scenarios.

Acknowledgment

This work was supported in part by the National Natural Science Foundation of China under Grant 62225107, 62371124, 61720106003, in part by the Fundamental Research Funds for the Central Universities 2242022k60002.

References

[1] C.-X. Wang et al., "On the Road to 6G: Visions, Requirements, Key Technologies, and Testbeds," *IEEE Communications Surveys & Tutorials*, vol. 25, no. 2, pp. 905-974, Second quarter 2023.

[2] Y. Gao et al., "Artificial Intelligence Enabled Joint Channel Estimation and Signal Detection for Massive MIMO Systems," in *Chinese Journal of Electronics*, vol. 35, no. 1, pp. 178-195, January 2026.

[3] M. Cui and L. Dai, "Channel Estimation for Extremely Large-Scale MIMO: Far-Field or Near-Field?" *IEEE Transactions on Communications*, vol. 70, no. 4, pp. 2663-2677, April 2022.

[4] Y. Zhang, X. Wu and C. You, "Fast Near-Field Beam Training for Extremely Large-Scale Array," *IEEE Wireless Communications Letters*, vol. 11, no. 12, pp. 2625-2629, Dec. 2022.

[5] M. Cui, L. Dai, Z. Wang, S. Zhou and N. Ge, "Near-Field Rainbow: Wideband Beam Training for XL-MIMO," *IEEE Transactions on Wireless Communications*, vol. 22, no. 6, pp. 3899-3912, June 2023.

[6] Y. Lu, Z. Zhang and L. Dai, "Hierarchical Beam Training for Extremely Large-Scale MIMO: From Far-Field to Near-Field," *IEEE Transactions on Communications*, vol. 72, no. 4, pp. 2247-2259, April 2024.

[7] J. Zhu, X. Ma, Z. Wang and Y. Huang, "Cluster-Based Low-Complexity Codebook Design for Hierarchical Beam Training in XL-MIMO," *2025 International Wireless Communications and Mobile Computing*, Abu Dhabi, United Arab Emirates, 2025, pp. 836-840.

[8] K. Chen, C. Qi, C. -X. Wang and G. Y. Li, "Beam Training and Tracking for Extremely Large-Scale MIMO Communications," *IEEE Transactions on Wireless Communications*, vol. 23, no. 5, pp. 5048-5062, May 2024.

[9] Z. Xiao, T. He, P. Xia and X. -G. Xia, "Hierarchical Codebook Design for Beamforming Training in Millimeter-Wave Communication," *IEEE Transactions on Wireless Communications*, vol. 15, no. 5, pp. 3380-3392, May 2016.

[10] J. Lee, G. -T. Gil and Y. H. Lee, "Channel Estimation via Orthogonal Matching Pursuit for Hybrid MIMO Systems in Millimeter Wave Communications," *IEEE Transactions on Communications*, vol. 64, no. 6, pp. 2370-2386, June 2016.