This article has been accepted for publication in IEEE Transactions on Wireless Communications. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TWC.2025.3575718

Latent Thompson Sampling-Based mmWave Receive Beam Measurement and Selection to Tackle User Orientation Changes and Mobility

Ashok Kumar Reddy Chavva, *Senior Member, IEEE* and Neelesh B. Mehta, *Fellow, IEEE*

***Abstract*—Beamforming enables millimeter-wave communications to achieve high data rates in 5G and beyond systems. However, accurate beam alignment entails a large training overhead. User device orientation changes and mobility can rapidly lead to beam misalignment and lower the data rate. They also make the beam gains a non-stationary random processes. We propose a comprehensive and novel approach called latent Thompson sampling-based beam selection (LTBS), which combines latent Thompson sampling to track the angle of arrival (AoA) as a latent state, receive beam subset selection based on the sampled AoA in a manner compliant with the 5G new radio standard, rate adaptation, and data beam selection based on predicted throughput. We propose two variants of LTBS that trade-off between complexity and accuracy in modeling millimeter-wave channels. The prior update and channel gain prediction in one of the variants are based on the realistic spatial channel model (SCM). We propose variations that employ windowing to also tackle lateral user mobility, which alters the AoA and the channel statistics. Our numerical results show that the proposed methods track the AoA in a manner robust to user orientation changes and provide higher average data rates compared to conventional and state-of-the-art learning-based beam selection methods.**

***Index Terms*—mmWave, 5G, latent Thompson sampling, beamforming, orientation change, angle-of-arrival, mobility**

I. Introduction

Millimeter-wave (mmWave) bands are essential for 5G and beyond systems to deliver high data rates. Narrow beams with large array gains are essential to overcome the severe propagation loss and attenuation from precipitation and atmospheric gases in these bands [1]. However, their narrow beamwidth requires accurate beam alignment and tracking [1]. Furthermore, a large number of transmit-receive beam pairs are needed to cover the full angular space in elevation and azimuth. For example, 16 to 64 transmit beams at the base station (BS) and 8 to 25 receive beams at the user equipment (UE), which corresponds to 128 to 1625 beam pairs, and a beamwidth of 5° to 20° are now typical in mmWave systems [2]–[4].

One of the difficult challenges faced in practice by mmWave beamforming is the rapid beam misalignment caused by changes in the orientation of the UE [5], [6]. The orientation of a handheld device can change at a rate as high as 110°/s [2] in non-gaming scenarios. Consequently, the UE needs to perform beam realignment and perhaps even sound all the possible beams frequently, which leads to a large training overhead. Another consequence of user orientation change is that the mean transmit-receive beam pair gain changes with time [7]. Thus, each beam pair's gain becomes a non-stationary random process. Movements by the UE, which we shall refer to as lateral mobility, cause even the angle of arrival (AoA) at the UE to vary. This again causes beam misalignment and a reduction in signal power and data rate. It also makes each beam pair's gain a non-stationary random process.

The training procedure in the 5G new radio (NR) standard for beam selection and alignment is as follows. The BS and the UE together determine the transmit-receive beam pair for data transmission. The BS transmits synchronization signal blocks (SSBs), which serve as reference signals, periodically [1], [8]. The SSBs are transmitted in an SSB burst, which has a duration of 5 ms. Different SSBs in a burst are transmitted using different BS transmit beams so that a UE can measure the beam gains from all the transmit beams of the BS to one of its receive beams. The SSB burst is transmitted with a periodicity that ranges from 5 ms to 160 ms.

For the UE to estimate gains of all the beam pairs, multiple SSB bursts are required – one for each UE receive beam. The UE selects its receive beam and the corresponding BS transmit beam periodically once every beam measurement cycle, which consists of multiple SSB bursts and slots for data transmission, and feeds back the transmit beam index to the BS. However, it does not need to report its selected receive beam to the BS. The UE also feeds back the channel quality information (CQI), which enables the BS to adjust its modulation and coding scheme (MCS) and data rate. In 5G NR, CQI is fed back more often than the transmit beam index [8]. The BS then sends the data in the next cycle on the selected transmit beam.

The significant time required for measuring the different transmit-receive beam pair gains implies that the measurements of the different beam pair gains are outdated by different extents by the time data transmission occurs. This outdatedness depends on the number of beam pairs measured, the order in which they are measured, and the rate at which the channel varies. Thus, user orientation changes and lateral mobility, which cause the channel to vary, and the training protocol have a significant impact on the data rates achieved by mmWave systems.

A. Related Literature on Beam Selection Methods

We classify the literature on beam selection methods into two categories, namely conventional and learning-based meth-

Ashok Kumar Reddy Chavva is with the Samsung Research Institute, Bangalore, India, and Neelesh B. Mehta is with the Department of Electrical Communication Engineering (ECE), Indian Institute of Science (IISc), Bangalore, India (emails: ashok.chavva@samsung.com, nbmehta@iisc.ac.in).

ods.

1) Conventional Methods: In exhaustive search, all transmit-receive beam pairs are sounded sequentially [1], [9]. However, this approach fails to utilize the information obtained from the measurements in previous cycles. Consequently, it takes longer for it to sound all beams, even if the previous measurements indicate that other beams are better. In hierarchical search, transmit beams are sounded over multiple rounds based on measurements obtained in the previous rounds and the beamwidths are successively reduced [10]. In [2], procedures based on predefined thresholds and beam power measurements are proposed to manage lateral mobility and changes in user orientation. In [11], matching theory is used to determine which transmit beams of the access points to assign to which UEs in a cell-free mmWave massive multiple-input multiple-output (MIMO) system. However, these methods choose the beam pair or the beam with the largest measured signal strength. They do not account for the measurements getting outdated.

In [5]–[7], orientation sensors on the UE are used to facilitate beam selection. In [7], a beam selection approach based on a statistical model of the spatial channel model (SCM) is presented. However, it assumes the AoA is known to the UE. In [12], channel sparsity and correlated beam gains are exploited to reduce the search space.

2) Learning-Based Methods: Learning-based multi-armed bandit (MAB) methods balance exploitation and exploration. They systematically explore different actions to discover potentially better options while simultaneously exploiting known good actions. MAB methods based on ϵ-greedy [13], upper confidence bound (UCB)-based [14], and Thompson sampling [15] have been proposed in the literature to select the beams. The ϵ-greedy method randomly selects an arm with probability ϵ while exploiting the arm with the highest reward with probability $1 - \epsilon$. However, choosing an appropriate, scenario-specific value of ϵ is challenging. The UCB-based method selects the arm with the highest UCB, which combines an estimated reward and a confidence interval that captures the uncertainty in that arm's reward.

The beam training procedure relies on the BS broadcasting reference signals and the UE reporting the BS transmit beam index. Additionally, the UE experiences orientation changes, whereas the BS does not. Therefore, beam pair selection at the BS must be implemented differently from that at the UE. We, therefore, discuss the BS-side and UE-side learning-based algorithms separately below.

a) BS Side: In [17], the unimodal structure of the mean received signal strength is exploited to eliminate suboptimal transmit beams at the BS. In [18], this is combined with hierarchical beamforming. In [19], [20], the BS uses an acknowledgment (ACK) or a negative ACK (NACK) obtained on each beam from the UE to learn the changes in beam quality over time. In [21], a learning-based beam grouping scheme is proposed to explore multiple directions concurrently and reduce the beam search time. In [22], a beam sampling algorithm is proposed that tracks the received signal strength to detect a change in the channel correlation and adaptively learn the best beam pair. In [15], Thompson sampling with Bernoulli rewards is employed. A hierarchical beam exploration method is proposed in [23], [24], but for the IEEE 802.11ad standard. However, these works assume that the UE has either omni-directional or wide beams; they do not consider beam selection at the UE. Hence, the problem of joint transmit-receive beam pair selection does not arise in this case. In [25], a learning-based deep Q-network is used to select a BS receive beam for an unmanned aerial vehicle (UAV). The UAV shares its location with the BS over a sub-6 GHz frequency. The BS then uses the UAV's location and uplink transmissions to select the beam pair.

b) UE Side: In [4], a contextual bandits-based method is proposed that uses the unimodal structure of the beam pattern to reduce the beam search time at the UE. In [14], a beam alignment algorithm based on the discounted UCB algorithm is proposed to handle non-stationary scenarios. The algorithm is run independently at the transmitter and receiver, with a signal-to-interference-and-noise ratio (SINR) threshold-based Bernoulli reward. In [26], a UCB-inspired UE beam set selection method is proposed. In [27], the data-driven long short-term memory (LSTM) learning method is used to predict which subset of beams will have the highest power. These beams are then selected for measurement.

Comments: Several important practical aspects of 5G NR are not modeled in the above works. For example, data and measurements happen in a time-interleaved fashion in 5G NR. However, [4], [17], [22] assume that pilot-based measurements happen first followed by data, while the measurement models in [14], [15], [18], [20], [21], [23] do not specify a pilot-based measurement model and are not practical. In [1], [4], [9]–[12], [15], [26], the time-evolution of the channel is not modeled and the channel is assumed to not change over a short time duration. And, [14] does not account for the measurements being outdated by different extents. However, with UE orientation changes and the timescale of transmission of the reference signals and the transmit beam feedback, this is no longer true. In [4], [14], [15], rewards based on Bernoulli outcomes are assumed. However, the Bernoulli distribution does not match the statistics of the real-valued beam gains or even their range. While a non-binary reward is considered in [26], it does not accurately represent the measurement outcome of a practical mmWave channel. UE-side aspects such as orientation changes and lateral mobility are not modeled in [4], [14], [15], [17]–[19], [21]–[23], [27].

While [7] accounts for the time evolution of the channel and considers SCM, it assumes that the AoA is known a priori to the UE. This is a major limitation, because even the AoA needs to be estimated by the UE. Its estimation is closely inter-twined with the beam selection process, which yields the measurements required to estimate the AoA, and requires an altogether different approach and performance evaluation.

The BS-side beam selection algorithms proposed in [17]–[19], [21]–[24] cannot be applied for UE receive beam selection due to asymmetry in the pilots transmitted by the BS and the UE. While the BS broadcasts periodic pilot bursts, this is not possible for the UE. In 5G NR, the UE does not need to feed back its receive beam to the BS. Therefore, this information is not available at the BS. Furthermore, orientation

TABLE I
COMPARISON OF RELATED LITERATURE ON MMWAVE BEAM SELECTION ('C' STANDS FOR CONVENTIONAL AND 'L' FOR LEARNING-BASED)

Reference	*Beam selection focus*	*UE orientation changes*	*Lateral mobility*	*Best K subset*	*Beam correlation*	*Channel non-stationarity*	*Time evolution*	*Method*
[7]	UE side	Yes	No	No	Yes	Yes	Yes	C
[2]	UE side	Yes	Yes	No	Yes	Yes	No	C
[12]	Joint BS-UE	No	No	Yes	Yes	No	No	C
[11]	BS side	No	No	No	No	No	No	C
[15]	BS side	No	Yes	No	No	Yes	No	L
[17]	BS side	No	No	No	Yes	No	No	L
[18]	BS side	No	No	No	Yes	No	Yes	L
[19]	BS side	No	Yes	No	No	Yes	Yes	L
[20]	BS side	No	No	No	No	Yes	Yes	L
[21]	BS side	No	No	No	No	Yes	No	L
[22]	BS side	No	Yes	No	No	Yes	No	L
[23]	BS side	No	No	No	Yes	No	No	L
[24]	BS side	No	No	No	Yes	No	No	L
[25]	BS side	No	Yes	No	No	No	Yes	L
[28]	BS side	No	Yes	No	No	No	Yes	L
[14]	UE side	No	No	No	No	Yes	Yes	L
[4]	UE side	No	No	No	Yes	No	No	L
[26]	UE side	No	No	Yes	No	No	No	L
[27]	UE side	No	Yes	Yes	Yes	No	Yes	L
This work	**UE side**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**L**

changes occur at the UE and not at the BS, which is static.

Table I provides a concise comparison of the literature on mmWave beam selection and alignment.

B. Focus and Contributions

We present a novel approach that innovatively combines receive beam subset selection for measuring transmit-receive beam pairs, latent Thompson sampling-based AoA estimation and tracking, domain-specific SCM-based prior, reward, and latent space sampling, and channel gain prediction-based beam selection for data. Our pilot, data, and feedback model is compliant with the 5G NR standard. We account for user orientation changes, lateral mobility, and the non-stationary time evolution of the mmWave channel. SCM has been adopted by the third generation partnership project (3GPP) for 5G. It provides a realistic, comprehensive mmWave channel model, unlike the simplistic models considered in [18], [23], [26]. It closely matches results from several measurement campaigns. It employs a geometric model for the AoA and angle of departure (AoD). It considers light-of-sight (LoS) and non-LoS (NLoS) paths, and scattering clusters and multiple paths per cluster [29].

We make the following contributions:

- We propose learning-based latent Thompson beam sampling (LTBS) to select the subset of receive beams to measure at the UE. It employs a probabilistic model for the AoA at the UE and tracks it as a latent state. In every beam measurement cycle, LTBS updates its prior distribution of the AoA based on the beam pairs measured by the UE. In turn, the receive beam subset to measure is updated based on the AoA estimate. We propose two variants, namely LTBS using beta-Bernoulli (LTBS-BB) conjugate pair probability density functions (PDFs) and LTBS using SCM (LTBS-SCM). LTBS-SCM incorporates a domain-specific reward and prior distribution that capture key physical attributes of the mmWave channel. This results in faster convergence and accurate beam selection even in a rapidly changing channel.
- In addition, we utilize the latent AoA estimate and the predicted beam pair gains at the times of data transmission to select the beam pair. Thus, the beam measurements being outdated by different extents is explicitly accounted for in our joint selection of the transmit-receive beam pair.
- We then extend the methods to address lateral mobility, which results in a time-varying AoA. We propose a windowing approach, which only uses recent measurements to update the prior and track the AoA in an agile manner.
- Our numerical results evaluate multiple performance metrics. They show that the proposed LTBS methods shortlist the best beam for measurement with a probability that exceeds 95% at all orientation change rates, while the corresponding probabilities of the benchmarks, which includes conventional as well as learning-based methods, can fall below 40%. The LTBS methods also track the AoA accurately. For example, at an orientation change rate of 60°/s, the standard deviation of the AoA estimation error is only 0.4° without lateral mobility and 2.1° with lateral mobility. These values are much smaller than the beamwidth. This combined with prediction-based data beam selection enables LTBS to achieve higher average data rates than all benchmarks.

Among the MAB methods, Thompson sampling employs a different Bayesian framework compared to UCB and ϵ-greedy [30]. It updates the prior based on the measurements obtained so far. Additionally, its actions are determined based on random variables drawn from the posterior distributions. As pointed out in [31], each posterior probability converges to the likelihood that the corresponding action maximizes reward, conditioned on the observed history. The numerical results in the classical MAB literature show that Thompson sampling achieves a lower mean regret per period compared to ϵ-greedy [31] and UCB [32]. The Bayesian framework also makes it easy to incorporate a latent state, as it can be included as an additional condition when the actions are selected using the prior. The practical application of Thompson sampling to mmWave beam selection using AoA as the latent state, and modeling the salient features of the 5G mmWave channel model and the beam selection procedure to effectively handle user orientation changes and lateral mobility is novel and significant. So is our adaption of non-stationary latent MAB to address AoA changes due to lateral mobility.

C. Organization and Notations

Section II presents the system model and SCM. In Section III, we propose the LTBS-BB and LTBS-SCM methods. In Section IV, we address lateral mobility. Section V contains our numerical results. Our conclusions follow in Section VI.

Notations: We denote the probability density function (PDF) of a random variable (RV) X by $f_X(\cdot)$. Similarly, the

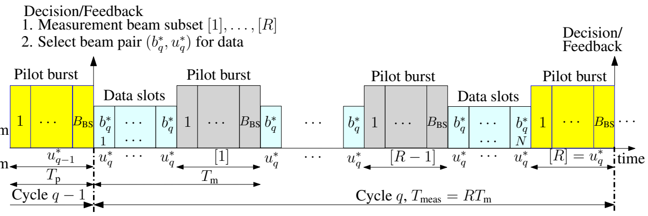

Fig. 1. Beam selection occurs once in a beam measurement cycle. Every pilot burst consists of $B_{\rm BS}$ pilot symbols, one from each of the transmit beams. Each cycle consists of N data slots that are interleaved with the pilot bursts.

conditional PDF conditioned on an event A are denoted by $f_X(\cdot|A)$. The expectation with respect to an RV X is denoted by $\mathbb{E}_X[\cdot]$ and the expectation conditioned on an event A by $\mathbb{E}_X[\cdot|A]$. The covariance of RVs X and Y is denoted by $\mathrm{Cov}(X,Y)$ and variance of RV X by $\mathrm{Var}(X)$. For a vector $\mathbf{x}$, $\mathbf{x}(i)$ denotes its i^{th} element, transpose by $\mathbf{x}^T$, and Hermitian transpose by $\mathbf{x}^\dagger$. The notation $X \sim \text{beta}(\alpha,\beta)$ implies that X is a beta RV with parameters α and β.

II. System Model

We consider a mmWave analog beamformed system. Such a system is practically appealing as it requires only one radio frequency chain. The BS is equipped with a uniform linear array (ULA) that consists of N_{tx} antennas. It transmits on one beam chosen from among B_{BS} fixed directional beams in the azimuth direction. Similarly, the UE is equipped with a ULA that consists of N_{rx} antennas. It can receive on one beam chosen from among B_{UE} fixed directional beams in the azimuth direction. Let $\mathcal{B}_{\text{BS}} = \{1,\dots,B_{\text{BS}}\}$ and $\mathcal{B}_{\text{UE}} = \{1,\dots,B_{\text{UE}}\}$ denote the set of transmit beams at the BS and the set of receive beams at the UE, respectively. The beam direction θ_b^{tx} for the b^{th} transmit beam is $\theta_b^{\text{tx}} = (b-1)\pi/B_{\text{BS}}$, for $1 \le b \le B_{\text{BS}}$, and the beam direction θ_u^{rx} for the u^{th} receive beam is $\theta_u^{\text{rx}} = (u-1)\pi/B_{\text{UE}}$, for $1 \le u \le B_{\text{UE}}$. The mean beamwidths of the transmit and receive beams are denoted by Δ^{tx} and Δ^{rx}, respectively. We shall refer to the tuple (b,u), where $b \in \mathcal{B}_{\text{BS}}$ and $u \in \mathcal{B}_{\text{UE}}$, as a beam pair.

A. Beam Measurement and Data Transmission Model

The beam measurements and data transmissions take place within a beam measurement cycle, which we shall refer to as a cycle henceforth. Each cycle q consists of three phases that are interleaved in time, namely beam measurements, beam selection, and data transmission. In the beam measurements phase, the BS sends multiple bursts of pilots from different transmit beams and the UE measures the strengths of the signals it receives on a subset of the receive beams. In the beam selection phase, the UE jointly selects the transmit-receive beam pair. It feeds back to the BS the transmit beam to use in the next cycle. In the data transmission phase, the BS transmits data on the serving transmit beam b_q^* and the UE receives it on the serving receive beam u_q^*. These phases are illustrated in Fig. 1. We describe them in detail below and set up the notation.

1) Beam Measurements: Each cycle consists of $R \le B_{\text{UE}}$ pilot bursts. R depends on the frequency with which the system permits transmit beam updates. In a pilot burst, the BS transmits pilot symbols from its B_{BS} beams one after the other in a burst of duration T_{p}, while the UE receives on one of its receive beams. The pilot bursts are spaced T_{m} apart in time. Thus, the duration T_{meas} of a cycle is RT_{m}.

The UE shortlists $R-1$ beams from its B_{UE} beams excluding the serving beam u_q^*. The shortlisted beams are denoted by $[1],\dots,[R-1]$, where $[k]$ is the index of the receive beam for the k^{th} pilot burst in the cycle. The serving beam is sounded last in the R^{th} pilot burst, i.e., $[R] = u_q^*$. This enables the UE to obtain fresh measurements for its serving receive beam, which has good odds of being reselected in a slowly-varying environment, and avoids the possibility of not measuring it in the initial cycles. Thus, the UE measures the beam pairs $(1,[1]),(2,[1]),\dots,(B_{\text{BS}},[1])$ in the first pilot burst, the beam pairs $(1,[2]),(2,[2]),\dots,(B_{\text{BS}},[2])$ in the second pilot burst, and so on up to the R^{th} pilot burst. Hence, in cycle q, the UE measures using its receive beams $[1],[2],\dots,[R-1],u_q^*$.

For a beam pair (b,u), let $T_{b,u}$ denote the most recent measurement time (in the current or previous cycles) and let the corresponding gain be $g_{b,u}(T_{b,u})$. Let $\mathbf{g} = [g_{1,1}(T_{1,1}),\ \dots, g_{b,u}(T_{b,u}),\ \dots, g_{B_{\text{BS}},B_{\text{UE}}}(T_{B_{\text{BS}},B_{\text{UE}}})]$. When the UE measures receive beam $[i]$, let b_i denote the BS transmit beam that gives the largest gain:

$$b_i = \underset{k \in \mathcal{B}_{\text{BS}}}{\arg\max} \left\{ g_{k,[i]}\left(T_{k,[i]}\right) \right\}. \tag{1}$$

Let $\mathbf{b}_q = (b_1,\dots,b_R)$. Thus, for the receive beams measured, $\mathbf{b}_q$ contains the corresponding best BS transmit beams.

2) Beam Subset Selection for Measurement and Beam Selection for Data: At the end of the q^{th} cycle, the UE does three things for the next cycle: 1) it selects the subset of receive beams to be measured, 2) it selects the BS transmit and UE receive beam pair (b_{q+1}^*, u_{q+1}^*) jointly for receiving data in cycle $q+1$, and 3) it feeds back the transmit beam index b_{q+1}^* to the BS. This reporting is done using the serving beam pair (b_q^*, u_q^*). Thus, the BS can switch its transmit beam every T_{meas} seconds.[1]

3) Data Transmission: It consists of N slots, each of duration T_{s}. These slots span the entire cycle, and are located in between pilot bursts. Slot n starts at time t_n. At the beginning of every slot, the UE feeds back to the BS the rate at which it can receive data in that slot. This model captures the fact that the BS can adapt its data rate in every slot in 5G NR [1], [8]. The beam gain variations within a slot are negligible because T_{s} is small compared to the coherence time of the channel. The effective data rate $B\left(g_{b_q^*,u_q^*}(t)\right)$ in bits/s/Hz on the selected beam pair after accounting for the time spent on the pilots is

$$B\left(g_{b_q^*,u_q^*}(t)\right) = \left(1 - \frac{T_{\text{p}}}{T_{\text{m}}}\right) \log_2\left(1 + \frac{P_{\text{tx}} g_{b_q^*,u_q^*}^2(t)}{\sigma^2}\right), \tag{2}$$

where P_{tx} is the transmit power and σ^2 is the noise variance.

[1] The beam selection feedback delay and the time required by the hardware to switch beams at the BS and the UE are negligible compared to the time required for sounding the beam pairs [1], [8].

We note that the above protocol applies equally as well to multiple UEs since each UE can use the reference signals broadcast by the BS. Each UE can run beam selection independently and simultaneously without additional overhead.

B. Spatial Channel Model

Let $\psi(t)$ be the orientation of the UE with respect to the antenna array at time t. $\psi(t)$ is known to the UE from its orientation sensors [2], [6]. This assumption is justified because the error in $\psi(t)$ is at most 0.1% [33]. The MIMO channel matrix $\mathbf{H}(t, \psi(t)) \in \mathbb{C}^{N_{\text{rx}} \times N_{\text{tx}}}$ between the BS and the UE at time t, which also depends on $\psi(t)$, is given by [7], [29], [34]

$$\mathbf{H}(t, \psi(t)) = \sqrt{\frac{K\Lambda}{K+1}} \mathbf{u}_{\text{rx}}(\theta^{\text{rx}}_{\text{LoS}} + \psi(t)) \mathbf{u}^{\dagger}_{\text{tx}}(\theta^{\text{tx}}_{\text{LoS}}) + \sqrt{\frac{\Lambda}{(K+1)L}} \sum_{c=1}^{C} \sum_{l=1}^{L} \alpha_{c,l}(t) \mathbf{u}_{\text{rx}}(\theta^{\text{rx}}_{c,l} + \psi(t)) \mathbf{u}^{\dagger}_{\text{tx}}(\theta^{\text{tx}}_{c,l}), \quad (3)$$

where C is the number of clusters, L is the number of paths per cluster, Λ is the path-loss, $\mathbf{u}_{\text{rx}}(\cdot)$ is the array response at the receiver, $\mathbf{u}_{\text{tx}}(\cdot)$ is the array response at the transmitter, and K is the Rician factor. And, $\theta^{\text{tx}}_{\text{LoS}}$ and $\theta^{\text{tx}}_{c,l}$ are the LoS AoD and AoD of the l^{th} path in the c^{th} cluster at the BS relative to its ULA, respectively. Similarly, $\theta^{\text{rx}}_{\text{LoS}}$ is the LoS AoA and $\theta^{\text{rx}}_{c,l}$ is the AoA of the l^{th} path in the c^{th} cluster at the UE; both are relative to the UE's ULA when the UE is at a reference orientation of 0°. Hence, after accounting for the UE orientation, the AoA angle is $\theta^{\text{rx}}_{\text{LoS}} + \psi(t)$. The array response vectors $\mathbf{u}_{\text{rx}}(\cdot)$ and $\mathbf{u}_{\text{tx}}(\cdot)$ are given by

$$\mathbf{u}_{\text{rx}}(\theta) = \frac{1}{\sqrt{N_{\text{rx}}}} \left[1, e^{-j2\pi\mu^{\text{rx}}(\theta)}, \ldots, e^{-j2\pi(N_{\text{rx}}-1)\mu^{\text{rx}}(\theta)}\right]^T, \quad (4)$$

$$\mathbf{u}_{\text{tx}}(\theta) = \frac{1}{\sqrt{N_{\text{tx}}}} \left[1, e^{-j2\pi\mu^{\text{tx}}(\theta)}, \ldots, e^{-j2\pi(N_{\text{tx}}-1)\mu^{\text{tx}}(\theta)}\right]^T, \quad (5)$$

where $\mu^{\text{tx}}(\theta) = d^{\text{tx}} \cos(\theta)/\lambda$, d^{tx} is the antenna spacing at the transmitter, and λ is the wavelength. Similarly, $\mu^{\text{rx}}(\theta) = d^{\text{rx}} \cos(\theta)/\lambda$ and d^{rx} is the antenna spacing at the receiver. Lastly, $\alpha_{c,l}(t) = \bar{\alpha}_{c,l} \exp(j2\pi f_D t \cos(\omega_{c,l}))$, where f_D is the maximum Doppler shift, $\omega_{c,l} = \theta^{\text{rx}}_{c,l} - \theta_v + \psi(t)$ is the resultant angle at which ray l from cluster c impinges on the receiver antenna array when the UE moves at an angle θ_v, and $\bar{\alpha}_{c,l}$ models small-scale fading. $\bar{\alpha}_{c,l}$ is a circularly symmetric complex Gaussian RV with zero mean and variance γ_c, which is the fraction of power in the c^{th} cluster. The system model is shown in Fig. 2.

1) SCM Statistical Parameters: The AoA $\theta^{\text{rx}}_{c,l}$ of path l of cluster c is a Gaussian RV that is wrapped over an interval of 2π radians. Its PDF $f^{\text{rx}}_c(\theta)$, for $-\pi < \theta \leq \pi$, is given by [34]

$$f^{\text{rx}}_c(\theta) = \frac{1}{\sqrt{2\pi}\sigma_{\text{AoA},c}} \sum_{\ell=-\infty}^{\infty} \exp\left(\frac{-(\theta + 2\pi\ell - \bar{\theta}_{\text{AoA},c})^2}{2\sigma^2_{\text{AoA},c}}\right). \quad (6)$$

With a mild abuse of terminology, we call $\bar{\theta}_{\text{AoA},c}$ and $\sigma_{AoA,c}$ the mean and the standard deviation of $\theta^{\text{rx}}_{c,l}$, respectively. The AoD $\theta^{\text{tx}}_{c,l}$ is a wrapped Gaussian RV with mean $\bar{\theta}_{\text{AoD},c}$ and standard deviation $\sigma_{\text{AoD},c}$. We denote its PDF by $f^{\text{tx}}_c(\theta)$. In SCM, $\sigma_{\text{AoA},c}$ and $\sigma_{AoD,c}$ are themselves exponential RVs with means ξ_{AoA} and ξ_{AoD}, respectively [35]. $\bar{\theta}_{\text{AoA},c}$, ξ_{AoA}, $\bar{\theta}_{\text{AoD},c}$, and ξ_{AoD} depend on the environment [35, Tbl. XI].

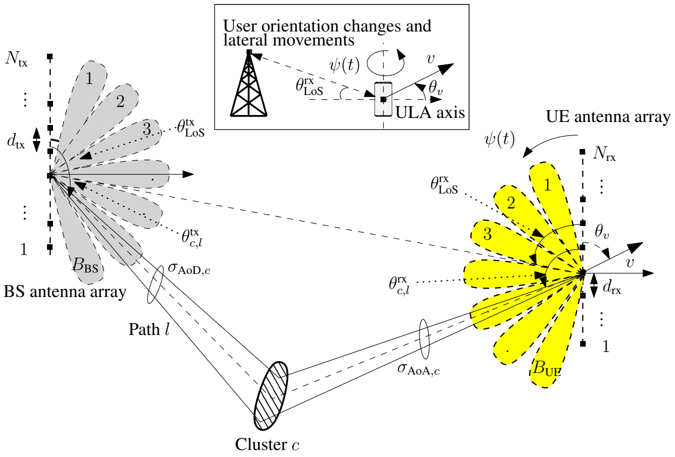

Fig. 2. A BS with N_{tx} antennas that can form one among B_{BS} beams and a UE with N_{rx} antennas that can form B_{UE} beams. Also shown is the user orientation $\psi(t)$, speed v, and clusters, AoD, and AoA of SCM.

2) Beam Gains: The beam gain $g_{b,u}(t)$ between the b^{th} transmit beam of the BS and the u^{th} receive beam of the UE at time t is given by

$$g_{b,u}(t) = \left|(\mathbf{u}_{\text{rx}}(\theta^{\text{rx}}_u)^{\dagger} \mathbf{H}(t, \psi(t)) \mathbf{u}_{\text{tx}}(\theta^{\text{tx}}_b)\right|, \quad (7)$$

where $\mathbf{u}_{\text{tx}}(\theta^{\text{tx}}_b)$ is the beamforming vector of the b^{th} transmit beam that points in the direction θ^{tx}_b and $\mathbf{u}_{\text{rx}}(\theta^{\text{rx}}_u)$ is the beamforming vector of the u^{th} receive beam that points in the direction θ^{rx}_u. The gains of two beam pairs can be correlated.

Note: $\psi(t)$ can model any orientation change trajectory; it need not depend on the speed v. As per SCM, the number of clusters, paths, AoAs, and AoDs change at a much slower rate than the the MIMO channel and beam gains [29], [34], [35]. We, therefore, assume that these remain constant. Lateral mobility causes the AoA to change with time, but at a much slower rate than $\psi(t)$.

C. Statistical Model for Time-Varying Beam-Pair Gain

Let $\rho_{b,u}(t, \theta^{\text{rx}}_{\text{LoS}})$ denote the power correlation coefficient between $g_{b,u}(t)$ and $g_{b,u}(t+\tau)$:

$$\rho_{b,u}(t, \theta^{\text{rx}}_{\text{LoS}}) \triangleq \frac{\mathbb{E}\left[g^2_{b,u}(t)\, g^2_{b,u}(t+\tau)\right] - \Omega_{b,u}(t, \theta^{\text{rx}}_{\text{LoS}})\, \Omega_{b,u}(t+\tau, \theta^{\text{rx}}_{\text{LoS}})}{\sqrt{\text{Var}(g^2_{b,u}(t)) \text{Var}(g^2_{b,u}(t+\tau))}}, \quad (8)$$

where $\Omega_{b,u}(t, \theta^{\text{rx}}_{\text{LoS}})$ is the mean channel power at measurement time t. The bivariate PDF $f_{g_{b,u}(t), g_{b,u}(t+\tau)}(r_1, r_2)$ of $g_{b,u}(t)$ and $g_{b,u}(t+\tau)$ is accurately characterized by the following modified Nakagami-m (MBN) model [7].

a) When $\rho_{b,u}\left(t,\theta_{LoS}^{rx}\right)\geq 0$: It is given by

$$
\begin{aligned}
&f_{g_{b,u}(t),g_{b,u}(t+\tau)}\left(r_1,r_2\right)\\
&=\frac{4m^{m+1}r_1^m r_2^m\left(\Omega_{b,u}\left(t,\theta_{\mathrm{LoS}}^{\mathrm{rx}}\right)\Omega_{b,u}\left(t+\tau,\theta_{\mathrm{LoS}}^{\mathrm{rx}}\right)\right)^{-\frac{m+1}{2}}}{\Gamma(m)\left[1-\rho_{b,u}\left(t,\theta_{\mathrm{LoS}}^{\mathrm{rx}}\right)\right]\left(\rho_{b,u}\left(t,\theta_{\mathrm{LoS}}^{\mathrm{rx}}\right)\right)^{\frac{m-1}{2}}}\\
&\times\exp\!\left(\frac{-m}{1-\rho_{b,u}\left(t,\theta_{\mathrm{LoS}}^{\mathrm{rx}}\right)}\left[\frac{r_1^2}{\Omega_{b,u}\left(t,\theta_{\mathrm{LoS}}^{\mathrm{rx}}\right)}+\frac{r_2^2}{\Omega_{b,u}\left(t+\tau,\theta_{\mathrm{LoS}}^{\mathrm{rx}}\right)}\right]\right)\\
&\times I_{m-1}\left(\frac{2mr_1r_2\sqrt{\rho_{b,u}\left(t,\theta_{\mathrm{LoS}}^{\mathrm{rx}}\right)}}{\sqrt{\Omega_{b,u}\left(t,\theta_{\mathrm{LoS}}^{\mathrm{rx}}\right))\Omega_{b,u}\left(t+\tau,\theta_{\mathrm{LoS}}^{\mathrm{rx}}\right)}\left[1-\rho_{b,u}\left(t,\theta_{\mathrm{LoS}}^{\mathrm{rx}}\right)\right]}\right),\\
&\qquad\qquad\text{for } r_1\geq 0, r_2\geq 0, \qquad (9)
\end{aligned}
$$

where m is the Nakagami parameter and $I_m\left(.\right)$ is the modified Bessel function of the first kind with order m [36, (9.6.19)]. The expressions for the MBN parameters $\Omega_{b,u}\left(t+\tau,\theta_{\mathrm{LoS}}^{\mathrm{rx}}\right)$, $\rho_{b,u}\left(t,\theta_{\mathrm{LoS}}^{\mathrm{rx}}\right)$, and m are given in Appendix B.

Therefore, the marginal PDF $f_{g_{b,u}(t)}\left(r\right)$ of $g_{b,u}\left(t\right)$ is $f_{g_{b,u}(t)}\left(r_1\right)=\int_0^\infty f_{g_{b,u}(t),g_{b,u}(t+\tau)}\left(r_1,r_2\right)dr_2$. Substituting for $f_{g_{b,u}(t),g_{b,u}(t+\tau)}\left(r_1,r_2\right)$ from (9) and using the identity $\int_0^\infty x^{\nu+1}e^{-\alpha x^2}I_\nu\left(\beta x\right)dx=\frac{\beta^\nu}{(2\alpha)^{\nu+1}}e^{-\frac{\beta^2}{4\alpha}}$, for $\alpha>0$ and $\nu>-1$ [37, (6.631.4)], we get

$$
f_{g_{b,u}(t)}\left(r_1\right)=\frac{2m^m r_1^{2m-1}}{\Gamma(m)\Omega_{b,u}^m\left(t,\theta_{\mathrm{LoS}}^{\mathrm{rx}}\right)}e^{\frac{-mr_1^2}{\Omega_{b,u}\left(t,\theta_{\mathrm{LoS}}^{\mathrm{rx}}\right)}},\ \text{for } r_1\geq 0. \qquad (10)
$$

The marginal PDF of $g_{b,u}\left(t+\tau\right)$ is the same as (10) except that $\Omega_{b,u}\left(t,\theta_{\mathrm{LoS}}^{\mathrm{rx}}\right)$ is replaced with $\Omega_{b,u}\left(t+\tau,\theta_{\mathrm{LoS}}^{\mathrm{rx}}\right)$.

b) When $\rho_{b,u}\left(t,\theta_{LoS}^{rx}\right)<0$: In this case, the bivariate PDF of $g_{b,u}\left(t\right)$ and $g_{b,u}\left(t+\tau\right)$ is given by [7]

$$
\begin{aligned}
&\tilde{f}_{g_{b,u}(t),g_{b,u}(t+\tau)}\left(r_1,r_2\right)=\frac{4m^{m+1}\left(a-r_1\right)^m r_2^m}{\zeta(m,2(2m-1))}\\
&\quad\times\frac{\left[\Omega_{b,u}\left(t,\theta_{\mathrm{LoS}}^{\mathrm{rx}}\right)\Omega_{b,u}\left(t+\tau,\theta_{\mathrm{LoS}}^{\mathrm{rx}}\right)\right]^{-\frac{m+1}{2}}}{\left(1-\left|\rho_{b,u}\left(t,\theta_{\mathrm{LoS}}^{\mathrm{rx}}\right)\right|\right)\left|\rho_{b,u}\left(t,\theta_{\mathrm{LoS}}^{\mathrm{rx}}\right)\right|^{\frac{m-1}{2}}}\\
&\times\exp\!\left(\frac{-m}{1-\left|\rho_{b,u}\left(t,\theta_{\mathrm{LoS}}^{\mathrm{rx}}\right)\right|}\left[\frac{\left(a-r_1\right)^2}{\Omega_{b,u}\left(t,\theta_{\mathrm{LoS}}^{\mathrm{rx}}\right)}+\frac{r_2^2}{\Omega_{b,u}\left(t+\tau,\theta_{\mathrm{LoS}}^{\mathrm{rx}}\right)}\right]\right)\\
&\times I_{m-1}\left(\frac{2m\sqrt{\left|\rho_{b,u}\left(t,\theta_{\mathrm{LoS}}^{\mathrm{rx}}\right)\right|}\left(a-r_1\right)r_2}{\sqrt{\Omega_{b,u}\left(t,\theta_{\mathrm{LoS}}^{\mathrm{rx}}\right)\Omega_{b,u}\left(t+\tau,\theta_{\mathrm{LoS}}^{\mathrm{rx}}\right)}\left(1-\left|\rho_{b,u}\left(t,\theta_{\mathrm{LoS}}^{\mathrm{rx}}\right)\right|\right)}\right),\\
&\qquad\qquad\text{for } 0\leq r_1\leq a, r_2\geq 0, \qquad (11)
\end{aligned}
$$

where $a=\sqrt{2(2m-1)\Omega_{b,u}\left(t,\theta_{\mathrm{LoS}}^{\mathrm{rx}}\right)/m}$ and $\zeta(\cdot,\cdot)$ is the incomplete gamma function [37, (8.350.1)]. We derive the marginal PDFs of $g_{b,u}\left(t+\tau\right)$ and $g_{b,u}\left(t\right)$ below.

i) Marginal PDF $f_{g_{b,u}(t+\tau)}\left(r_2\right)$ of $g_{b,u}\left(t+\tau\right)$: It is given by

$$
f_{g_{b,u}(t+\tau)}\left(r_2\right)=\int_0^a\tilde{f}_{g_{b,u}(t),g_{b,u}(t+\tau)}\left(r_1,r_2\right)dr_1. \qquad (12)
$$

Using the infinite series expansion $I_\nu\left(z\right)=\left(\frac{z}{2}\right)^\nu\sum_{k=0}^\infty\frac{\left(\frac{z^2}{4}\right)^k}{k!\,\Gamma(\nu+k+1)}$ [36, (9.6.10)] in (11), we can show that

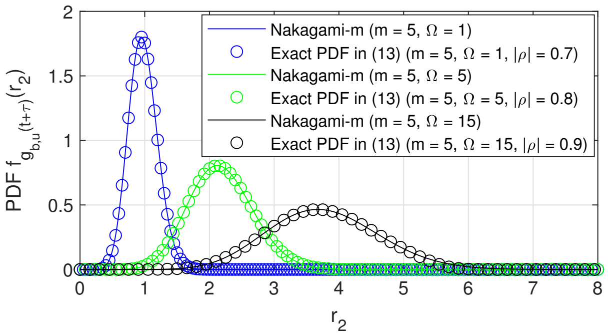

Fig. 3. Comparison of the exact PDF in (13) and the Nakagami-m PDF for different values of ρ, Ω, and m.

$$
\begin{aligned}
f_{g_{b,u}(t+\tau)}\left(r_2\right)&=\frac{2m^m r_2^{2m-1}}{\zeta(m,2(2m-1))(\Omega_{b,u}\left(t+\tau,\theta_{\mathrm{LoS}}^{\mathrm{rx}}\right))^m}\\
&\times\exp\left(\frac{-mr_2^2}{\left(1-\left|\rho_{b,u}\left(t,\theta_{\mathrm{LoS}}^{\mathrm{rx}}\right)\right|\right)\Omega_{b,u}\left(t+\tau,\theta_{\mathrm{LoS}}^{\mathrm{rx}}\right)}\right)\\
&\times\sum_{k=0}^\infty r_2^{2k}\left(\frac{m\left|\rho_{b,u}\left(t,\theta_{\mathrm{LoS}}^{\mathrm{rx}}\right)\right|}{\left(1-\left|\rho_{b,u}\left(t,\theta_{\mathrm{LoS}}^{\mathrm{rx}}\right)\right|\right)\Omega_{b,u}\left(t+\tau,\theta_{\mathrm{LoS}}^{\mathrm{rx}}\right)}\right)^k\\
&\times\frac{\zeta\left(m+k,\frac{2(2m-1)}{\left(1-\left|\rho_{b,u}\left(t,\theta_{\mathrm{LoS}}^{\mathrm{rx}}\right)\right|\right)}\right)}{k!\,\Gamma(m+k)},\ r_2\geq 0. \qquad (13)
\end{aligned}
$$

While this expression is exact, its infinite series form makes it computationally expensive, which is a problem since it is needed later to compute the expected reward and to update the prior. Motivated by the fact that the bivariate PDF in (11) is derived by a series of transformations of the bivariate Nakagami-m PDF in (9) that consist of a normalization, a rotation, and a rescaling, we instead approximate (13) by the Nakagami-m PDF in (10).[2] We assess the accuracy of the approximation below.

Fig. 3 plots the marginal PDFs in (10) and (13) for different values of m, $\Omega_{b,u}^m\left(t+\tau,\theta_{\mathrm{LoS}}^{\mathrm{rx}}\right)$ (referred as Ω in the plots), and $\rho_{b,u}\left(t,\theta_{\mathrm{LoS}}^{\mathrm{rx}}\right)$ (referred to as ρ in the plots). We observe that these two PDFs are indistinguishable from each other for all the parameter combinations considered. To better understand and quantify the accuracy of the approximation, we also compare these two marginal PDFs using the Kullback–Leibler (KL) divergence in Appendix C. We observe that the KL divergence is small (below 0.01) for $|\rho|\leq 0.9$. Thus, the Nakagami-m PDF is a good approximation for the PDF of $g_{b,u}\left(t+\tau\right)$.

ii) Marginal PDF $f_{g_{b,u}(t)}\left(r_1\right)$ of $g_{b,u}\left(t\right)$: It is given by

$$
f_{g_{b,u}(t)}\left(r_1\right)=\int_0^\infty\tilde{f}_{g_{b,u}\left(t,\theta_{\mathrm{LoS}}^{\mathrm{rx}}\right),g_{b,u}\left(t+\tau,\theta_{\mathrm{LoS}}^{\mathrm{rx}}\right)}\left(r_1,r_2\right)dr_2. \qquad (14)
$$

Substituting for $\tilde{f}_{g_{b,u}(t),g_{b,u}(t+\tau)}\left(r_1,r_2\right)$ from (11) and using the identity $\int_0^\infty x^{\nu+1}e^{-\alpha x^2}I_\nu\left(\beta x\right)dx=\frac{\beta^\nu}{(2\alpha)^{\nu+1}}e^{-\frac{\beta^2}{4\alpha}}$, for

[2] When $\rho_{b,u}\left(t,\theta_{\mathrm{LoS}}^{\mathrm{rx}}\right)=0$, the approximation is exact since (13) can be shown to simplify to (10).

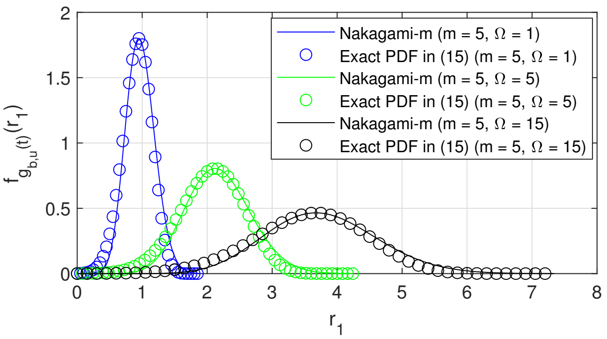

Fig. 4. Comparison of the exact PDF in (15) and the Nakagami-m PDF for different values of Ω and m.

$\alpha > 0$ and $\nu > -1$ [37, (6.631.4)], we get the following closed-form expression:

$$f_{g_{b,u}(t)}\left(r_1\right) = \frac{2m^m (a-r_1)^{2m-1}}{\zeta(m, 2(2m-1))\Omega_{b,u}^m\left(t, \theta_{\text{LoS}}^{\text{rx}}\right)} \times \exp\left(\frac{-m(a-r_1)^2}{\Omega_{b,u}\left(t, \theta_{\text{LoS}}^{\text{rx}}\right)}\right), \text{ for } 0 < r_1 \le a. \quad (15)$$

Note that the marginal PDF is not a function of $\rho_{b,u}\left(t, \theta_{\text{LoS}}^{\text{rx}}\right)$, unlike (13). We again approximate it by the same Nakagami-m PDF to avoid having different reward calculations for times t and $t+\tau$. We assess the accuracy of the approximation below.

Fig. 4 plots the PDFs in (10) and (15) for different values of m and $\Omega_{b,u}^m\left(t, \theta_{\text{LoS}}^{\text{rx}}\right)$ (referred as Ω in the plots). We observe that these two PDFs are very close to each other for all the parameter combinations considered. We also compare these two marginal PDFs using KL divergence in Appendix C. It is less than 0.1 for all parameter combinations considered. Hence, the Nakagami-m PDF can be used as the marginal PDF in all cases.

III. Measured Beam Subset Selection Through Latent Thompson Sampling

We first describe the general LTBS method. Then, we present its two variants. The method goes hand-in-hand with data beam selection, which we specify in Section III-C.

Latent Thompson sampling extends Thompson sampling by incorporating a hidden or latent state, which will be the AoA in our formulation. Over time, as more measurements become available to update the latent state, the prior distribution converges to the true, but unknown, underlying AoA ϕ^*, enabling better beam selection in time-varying channels. The learning agent resides at the UE.

We define the action space, context, latent space, reward, and prior distribution as follows:

- *Action Space* $\mathcal{A}$: An action $\mathbf{a}_q \in \mathcal{A}$ represents the vector of receive beams $\left([1], \ldots, [R-1], u_q^*\right)$ selected for measurement in the cycle $q \in \mathbb{N}$. The i^{th} element of $\mathbf{a}_q$ is the receive beam that is used to receive the i^{th} pilot burst in the cycle q. This beam points in the direction $\theta_{[i]}^{\text{rx}}$. The R^{th} element of $\mathbf{a}_q$ is the serving receive beam u_q^*. Thus, $\mathcal{A}$ contains all vectors of length R whose elements are drawn without repetition from $\mathcal{B}_{\text{UE}}$. The order of elements in $\mathbf{a}_q$ determines when a receive beam is used and the extent by which each measurement is outdated.
- *Context* $\mathbf{x}_q$: The context $\mathbf{x}_q \in \mathbb{R}^{qN \times 1}$ in cycle q is the sequence of user orientations from slot 1 of first cycle to the last (N^{th}) slot in the q^{th} cycle.
- *Latent Space* $\mathcal{D}$ *and State*: The latent space is $\mathcal{D} = \{1, \ldots, D\}$. It contains the possible AoAs of the LoS path at the UE. The latent state $d \in \mathcal{D}$ implies that the AoA ϕ_d is $\frac{2\pi(d-1)}{D}$. Here, the mean AoA has been discretized into D values. In general, to track the AoA accurately, we need $\frac{2\pi}{D} \ll \Delta^{\text{rx}}$, i.e., $D \gg \frac{2\pi}{\Delta^{\text{rx}}}$. Let $\tilde{\phi}_q$ be the AoA estimate in cycle q.
- *Reward*: Let $r_{q,[i]} \in \mathbb{R}$ be the reward for receive beam $[i]$ in cycle q. Let $\mathbf{r}_q = \left(r_{q,[1]}, \ldots, r_{q,[R]}\right)$. We denote its probability mass function conditioned on the action, context, and latent state by $\Pr\left(\mathbf{r}_q | \mathbf{a}_q, \mathbf{x}_q, \tilde{\phi}_q\right)$ for a discrete reward, and its PDF by $f\left(\mathbf{r}_q | \mathbf{a}_q, \mathbf{x}_q, \tilde{\phi}_q\right)$ for a continuous reward.
- *Prior Distribution*: We call the PDF of the states $(\phi_1, \ldots, \phi_D)$ at the start of cycle q as the prior distribution. We represent it with a vector of parameters $\mathbf{p}_q$.

We set up the prior update, AoA sampling, feedback and data beam pair selections, and action in cycle q as follows:

- *Prior Update*: The prior distribution is updated based on the rewards observed by the agent in the cycle and becomes the prior for the next cycle. The complexity of the prior update depends on the prior and the reward. We specify it below when we discuss the two variants.
- *AoA Sampling*: The agent samples the AoA $\tilde{\phi}_q$ from the distribution $\mathbf{p}_q$. This has a computational complexity of $\mathcal{O}(D)$.
- *Feedback and Data Beam Pair Selection*: We describe the method for selection of the beam pair (b_{q+1}^*, u_{q+1}^*) for data later in Section III-C. It is a function of $\tilde{\phi}_q$ and $\mathbf{x}_q$. The UE feeds back the transmit beam index b_{q+1}^* to the BS and uses the receive beam u_{q+1}^* in cycle $q+1$.
- *Action*: Let
$$\delta_k = \mathbb{E}\left[r_{q,k} | \tilde{\phi}_q, \mathbf{x}_q\right], \text{ for } k \in \mathcal{B}_{\text{UE}} \setminus \left\{u_{q+1}^*\right\}, \quad (16)$$
denote the expected reward from using receive beam k, except the serving one, given the sampled AoA and context. We sort $\delta_1, \delta_2, \ldots$ in the descending order. Then, $\mathbf{a}_{q+1}(1)$ is set as the receive beam with the largest expected reward, $\mathbf{a}_{q+1}(2)$ is the receive beam with the second largest reward, and so on. Thus,
$$\delta_{\mathbf{a}_{q+1}(1)} \ge \delta_{\mathbf{a}_{q+1}(2)} \ge \cdots \ge \delta_{\mathbf{a}_{q+1}(R-1)}. \quad (17)$$
Lastly, the R^{th} beam is the serving receive beam: $\mathbf{a}_{q+1}(R) = u_{q+1}^*$. In effect, the action maximizes $\sum_{i=1}^{R-1} \mathbb{E}\left[r_{q,i} | \tilde{\phi}_q, \mathbf{x}_q\right]$ and, in addition, sounds the serving beam.
We derive the expression for $\mathbb{E}\left[r_{q,u} | \tilde{\phi}_q, \mathbf{x}_q\right]$ when we specify the LTBS variants. The complexity of this step, which involves sorting B_{UE} elements is $\mathcal{O}(B_{\text{UE}} \log(B_{\text{UE}}))$.

Fig. 5 presents a flow chart of the various steps in LTBS. We now present two variants of the above general method, namely

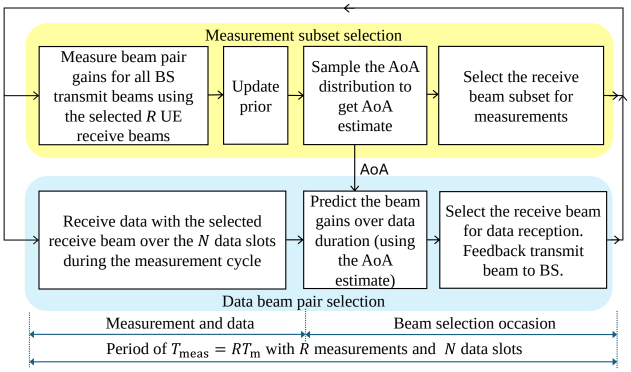
Measurement subset selection
Measure beam pair gains for all BS transmit beams using the selected R UE receive beams
Update prior
Sample the AoA distribution to get AoA estimate
Select the receive beam subset for measurements
AoA
Receive data with the selected receive beam over the N data slots during the measurement cycle
Predict the beam gains over data duration (using the AoA estimate)
Select the receive beam for data reception. Feedback transmit beam to BS.
Data beam pair selection
Measurement and data
Beam selection occasion
Period of $T_{\text{meas}} = RT_{\text{m}}$ with R measurements and N data slots

Fig. 5. Flowchart of measured beam subset selection, data beam selection and reporting.

LTBS-BB and LTBS-SCM. In LTBS-BB, a binary reward is defined and the conjugate beta PDF prior is used for it. On the other hand, LTBS-SCM employs the beam gain as the reward and updates the prior based on SCM.

A. LTBS-BB

We define the reward and prior in cycle q as follows:

- *Reward*: The $\mathbf{a}_q(i)^{\text{th}}$ element $r_{q,\mathbf{a}_q(i)}$ of the reward is

$$r_{q,\mathbf{a}_q(i)} = \begin{cases} 1, & \text{if } g^2_{\mathbf{b}_q(i),\mathbf{a}_q(i)}\left(T_{\mathbf{b}_q(i),\mathbf{a}_q(i)}\right) \geq \tau_{\text{BB}}, \\ 0, & \text{otherwise}, \end{cases} \quad (18)$$

  where τ_{BB} is a threshold. A reward of 1 indicates that the receive beam $[i]$ is strong when paired with at least one transmit beam. Thus, it is worth measuring. It also implicitly implies that the beam is aligned with the AoA.
- *Prior and AoA Sampling*: Let (α_d, β_d) denote the parameters that define the beta PDF for the latent state $d \in \mathcal{D}$. At $q = 1$, we set $(\alpha_d, \beta_d) = (1, 1), \forall d \in \mathcal{D}$. We refer to the vector of parameters $((\alpha_1, \beta_1), \ldots, (\alpha_D, \beta_D))$, which define the prior, itself as the prior $\mathbf{p}_q$.
  For each latent state d, we draw a sample $s_{q,d} \sim \text{beta}(\alpha_d, \beta_d)$. The sampled latent state $\tilde{d}$ is then the one with the largest sample value and is given by

$$\tilde{d} = \arg\max_{d \in \mathcal{D}} \left\{ s_{q,d} \right\}, \quad (19)$$

  and the sampled AoA $\tilde{\phi}_q$ is $\phi_{\tilde{d}}$.
- *Action*: The action $\mathbf{a}_{q+1}$ is computed as per (16) and (17). Using the PDF of $g_{\mathbf{b}_q(i),\mathbf{a}_q(i)}(t)$ in (10), we get

$$\mathbb{E}\left[r_{q,\mathbf{a}_q(i)} | \tilde{\phi}_q, \mathbf{x}_q\right] = \frac{2m^m}{\Gamma(m)\Omega^m_{\mathbf{b}_q(i),\mathbf{a}_q(i)}\left(t, \tilde{\phi}_q\right)} \times \int_{\sqrt{\tau_{\text{BB}}}}^{\infty} r^{2m-1} \exp\left(\frac{-m}{\Omega_{\mathbf{b}_q(i),\mathbf{a}_q(i)}\left(t, \tilde{\phi}_q\right)} r^2\right) dr. \quad (20)$$

  Using the identity $\int_u^\infty x^\alpha \exp\left(-\beta x^n\right) = \frac{\Gamma(\nu, \beta u^n)}{n\beta^\nu}$ [37, 3.381(9)], where $\nu = \frac{\alpha+1}{n}$ and $\Gamma(.,.)$ is the upper incomplete gamma function, (20) simplifies to

$$\mathbb{E}\left[r_{q,\mathbf{a}_q(i)} | \tilde{\phi}_q, \mathbf{x}_q\right] = \frac{\Gamma\left(m, \frac{m\tau_{\text{BB}}}{\Omega_{\mathbf{b}_q(i),\mathbf{a}_q(i)}\left(t, \tilde{\phi}_q\right)}\right)}{\Gamma(m)}. \quad (21)$$

- *Prior Update*: Let $\mathcal{Q}_{q,\mathbf{a}_q(i)}$ denote the set of AoA states in $\mathcal{D}$ covered by the beam $\mathbf{a}_q(i)$ in cycle q. It is given by

$$\mathcal{Q}_{q,\mathbf{a}_q(i)} = \left\{ d \in \mathcal{D} | \phi_d \in \left[\theta^{\text{rx}}_{\mathbf{a}_q(i)} - \psi\left(T_{\mathbf{b}_q(i),\mathbf{a}_q(i)}\right) - \frac{\Delta^{\text{rx}}}{2}, \right.\right. \\ \left.\left. \theta^{\text{rx}}_{\mathbf{a}_q(i)} - \psi\left(T_{\mathbf{b}_q(i),\mathbf{a}_q(i)}\right) + \frac{\Delta^{\text{rx}}}{2} \right) \right\}. \quad (22)$$

  We assume that the probability distributions of AoA states are independent. Therefore, for the receive beam $\mathbf{a}_q(i)$, we update the priors for each AoA index $d \in \mathcal{Q}_{q,\mathbf{a}_q(i)}$ as:

$$\alpha_d = \alpha_d + r_{q,\mathbf{a}_q(i)}, \quad (23)$$

$$\beta_d = \beta_d + 1 - r_{q,\mathbf{a}_q(i)}. \quad (24)$$

  This update rule follows since the beta PDF is the conjugate of the Bernoulli distribution. Note that an AoA's prior can be updated multiple times if it is covered by multiple beams. The priors of AoAs not covered by any measured beam in the cycle are not updated. The prior update above involves simple additions. Its complexity is $\mathcal{O}(D)$.

The pseudo-code of LTBS-BB is given in Algorithm 1.

Explanation: The initialization of the prior distribution for each angle in $\mathcal{D}$ ensures that all angles are equally likely initially. In cycle q, the UE measures the R selected beams, with the R^{th} beam being the serving receive beam. At the end of cycle q, for each measured receive beam, the UE computes the reward and updates the priors for all AoAs covered by the beam. Next, the UE samples the AoA using the beta PDFs for the D possible AoAs. The direction $\tilde{\phi}_q$ with the highest sampled value is selected. Next, the UE selects $R - 1$ receive beams with the largest expected rewards to be measured in the next cycle given the AoA sample and the context. The R^{th} receive beam is the serving beam u^*_q.

B. LTBS-SCM

In LTBS-SCM, the reward and prior are as follows:

- *Reward*: Let

$$\mathcal{C}_q = \left\{ (\mathbf{b}_q(1), \mathbf{a}_q(1)), \ldots, (\mathbf{b}_q(R), \mathbf{a}_q(R)) \right\}. \quad (25)$$

  It contains the R measured receive beams and the corresponding transmit beams in $\mathbf{b}_q$. For the i^{th} beam pair $(\mathbf{b}_q(i), \mathbf{a}_q(i)) \in \mathcal{C}_q$, the reward $r_{q,i}$ is defined as

$$r_{q,\mathbf{a}_q(i)} \triangleq \max\{ g^2_{\mathbf{b}_q(i),\mathbf{a}_q(i)}\left(T_{\mathbf{b}_q(i),\mathbf{a}_q(i)}\right), \tau_{\text{SCM}} \}, \quad (26)$$

  where τ_{SCM} is a cap. In practice, τ_{SCM} is determined by the minimum signal-to-noise ratio (SNR) value at which the highest-rate MCS permitted in the standard meets the block error rate target. This reward allows the UE to distinguish between two receive beams with different receive power levels and, thus, different rates.

Algorithm 1 LTBS-BB

```
**Initialization**: At $q = 1$, set the prior $(\alpha_d, \beta_d) = (1, 1), \forall d \in \mathcal{D}$, $\mathbf{a}_1$ by sampling uniformly from $\mathcal{B}_{\text{UE}}$ with out repetition, and orientation change context $\mathbf{x}_1$.
**for** $q \leftarrow 1, 2, \ldots$ **do**
    #measurements and rewards:
    Measure the beams in $\mathbf{a}_q$ in order and obtain reward $\mathbf{r}_q$
    #update latent state priors:
    **for** $i \leftarrow 1, 2, \ldots, R$ **do**
        $\alpha_d \leftarrow \alpha_d + r_{q,i}, \forall d \in \mathcal{Q}_{q,\mathbf{a}_q(i)}$
        $\beta_d \leftarrow \beta_d + 1 - r_{q,i}, \forall d \in \mathcal{Q}_{q,\mathbf{a}_q(i)}$
    **end for**
    #sample state model and select AoA:
    Sample $s_{q,d} \sim \text{beta}(\alpha_d, \beta_d), \forall d \in \mathcal{D}$
    $\tilde{d} \leftarrow \arg\max_{d \in \mathcal{D}} \{s_{q,d}\}$. Set $\tilde{\phi}_q = \phi_{\tilde{d}}$
    #select data beam for cycle $q + 1$:
    Select the data beam $u^*_{q+1}$ given $\mathbf{g}$, $\tilde{\phi}_q$, and $\mathbf{x}_q$ based on the method described in Section III-C
    #select measurement subset for cycle $q + 1$:
    Choose the receive beam array $\mathbf{a}_{q+1}$ as per (17) and (21)
**end for**
```

- *Prior and AoA Sampling*: Here, the prior is a multinomial PDF. We denote it by $\mathbf{p}_q = (P_{1,q}, \ldots, P_{D,q})$, where $P_{d,q}$ is the probability for selecting the state d in cycle q. For $q = 1$, we set $\mathbf{p}_1 = \left(\frac{1}{D}, \ldots, \frac{1}{D}\right)$ so that all states are equi-probable initially. The sampled AoA latent state is then given by

$$\tilde{d} = \text{mnrnd}(\mathbf{p}_q), \tag{27}$$

$$\tilde{\phi}_q = \phi_{\tilde{d}}, \tag{28}$$

where $\text{mnrnd}(\mathbf{p}_q)$ represents a sample drawn from the multinomial PDF $\mathbf{p}_q$.

- *Action*: The action $\mathbf{a}_{q+1}$ is computed as per (16) and (17). For the reward in (26), $\mathbb{E}\left[r_{q,\mathbf{a}_q(i)} | \tilde{\phi}_q, \mathbf{x}_q\right]$ is obtained by using the PDF of $g_{\mathbf{b}_q(i),\mathbf{a}_q(i)}(t)$ in (10). It is given by

$$\begin{aligned}\mathbb{E}\left[r_{q,\mathbf{a}_q(i)} | \tilde{\phi}_q, \mathbf{x}_q\right] &= \frac{2m^m}{\Gamma(m)\Omega^m_{\mathbf{b}_q(i),\mathbf{a}_q(i)}\left(t, \tilde{\phi}_q\right)} \\ &\times \left[\int_0^{\sqrt{\tau_{\text{SCM}}}} r^2 r^{2m-1} \exp\left(\frac{-m}{\Omega_{\mathbf{b}_q(i),\mathbf{a}_q(i)}\left(t, \tilde{\phi}_q\right)} r^2\right) dr,\right. \\ &\left.+ \int_{\sqrt{\tau_{\text{SCM}}}}^{\infty} \tau_{\text{SCM}} r^{2m-1} \exp\left(\frac{-m}{\Omega_{\mathbf{b}_q(i),\mathbf{a}_q(i)}\left(t, \tilde{\phi}_q\right)} r^2\right) dr\right].\end{aligned} \tag{29}$$

To simplify the first integration term in (29), we use the identity $\int_0^u x^\alpha \exp(-\beta x^n) = \frac{\gamma(\nu, \beta u^n)}{n\beta^\nu}$ [37, 8.350(1)], where $\nu = \frac{\alpha+1}{n}$, $\gamma(.,.)$ is the lower incomplete gamma function, and $\gamma(m, a) = \Gamma(m) - \Gamma(m, a)$. The second integration term can be simplified in a manner similar to (20). With these identities, we get

$$\begin{aligned}\mathbb{E}\left[r_{q,\mathbf{a}_q(i)} | \tilde{\phi}_q, \mathbf{x}_q\right] &= \frac{\tau_{\text{SCM}}\Gamma\left(m, \frac{m\tau_{\text{SCM}}}{\Omega_{\mathbf{b}_q(i),\mathbf{a}_q(i)}\left(t, \tilde{\phi}_q\right)}\right)}{\Gamma(m)} \\ +\Omega_{\mathbf{b}_q(i),\mathbf{a}_q(i)}\left(t, \tilde{\phi}_q\right) &\left[1 - \frac{\Gamma\left(m+1, \frac{m\tau_{\text{SCM}}}{\Omega_{\mathbf{b}_q(i),\mathbf{a}_q(i)}\left(t, \tilde{\phi}_q\right)}\right)}{\Gamma(m+1)}\right].\end{aligned} \tag{30}$$

When τ_{SCM} is large, it can be shown that

$$\mathbb{E}\left[r_{q,\mathbf{a}_q(i)} | \tilde{\phi}_q, \mathbf{x}_q\right] \approx \Omega_{\mathbf{b}_q(i),\mathbf{a}_q(i)}\left(t, \tilde{\phi}_q\right). \tag{31}$$

- *Prior Update*: Using the Bayes' theorem, $P_{d,q+1}$, $\forall d \in \mathcal{D}$, can be written as

$$P_{d,q+1} = \frac{f(\mathbf{r}_q | \phi_d)\, P_{d,q}}{\sum_{d'=1}^{D} f(\mathbf{r}_q | \phi_{d'})\, P_{d',q}}. \tag{32}$$

The gains of different beams are mutually independent because the rays from each cluster combine with different phases at the different antennas [29]. Hence,

$$f(\mathbf{r}_q | \phi_d) = \prod_{i=1}^{R} f(r_{q,i} | \phi_d) = \prod_{(b,u) \in \mathcal{C}_q} f\left(g^2_{b,u}(T_{b,u}) | \phi_d\right).$$

When τ_{SCM} is large, $r_{q,i} \approx g^2_{\mathbf{b}_q(i),\mathbf{a}_q(i)}\left(T_{\mathbf{b}_q(i),\mathbf{a}_q(i)}\right)$. Substituting the expression in (10), we get

$$\begin{aligned}f(\mathbf{r}_q | \phi_d) = \prod_{(b,u) \in \mathcal{C}_q} &\left[\frac{2m^m \left(g_{b,u}(T_{b,u})\right)^{2m-1}}{\Gamma(m)\Omega^m_{b,u}(t, \phi_d)}\right. \\ &\left.\times \exp\left(\frac{-m g^2_{b,u}(T_{b,u})}{\Omega_{b,u}(t, \phi_d)}\right)\right].\end{aligned} \tag{33}$$

Substituting (33) in (32), we obtain

$$\begin{aligned}&P_{d,q+1} = P_{d,q} \\ &\times \frac{\prod_{(b,u) \in \mathcal{C}_q} \left[\exp\left(\frac{-m g^2_{b,u}(T_{b,u})}{\Omega_{b,u}(t,\phi_d)}\right) \frac{1}{\Omega^m_{b,u}(t,\phi_d)}\right]}{\sum_{i=1}^{D} P_{i,q} \prod_{(b,u) \in \mathcal{C}_q} \left[\exp\left(\frac{-m g^2_{b,u}(T_{b,u})}{\Omega_{b,u}(t,\phi_i)}\right) \frac{1}{\Omega^m_{b,u}(t,\phi_i)}\right]}.\end{aligned} \tag{34}$$

Unlike LTBS-BB, the prior update above is more computationally involved. It involves evaluating the reward PDF R times for all the D AoAs. Its complexity is $\mathcal{O}(DR)$.

The pseudo-code of LTBS-SCM is given in Algorithm 2.

C. Data Beam Selection

The data beam pair for cycle $q + 1$ is selected at the beam selection occasion at the end of cycle q. We choose it as the beam pair that maximizes the predicted average data rate over the N data slots in cycle $q + 1$ given the beam pair measurements and sampled AoA $\tilde{\phi}_q$. Thus,

$$\begin{aligned}\left(b^*_{q+1}, u^*_{q+1}\right) = \underset{b \in \mathcal{B}_{\text{BS}}, u \in \mathcal{B}_{\text{UE}}}{\arg\max} &\left\{\frac{1}{N}\right. \\ &\left.\times \sum_{k=1}^{N} \mathbb{E}_{g_{b,u}(t_k)}\left[B\left(g_{b,u}(t_k)\right) | g_{b,u}(T_{b,u}), \tilde{\phi}_q\right]\right\}.\end{aligned} \tag{35}$$

Algorithm 2 LTBS-SCM

```
Initialization: At $q = 1$, set $\mathbf{p}_1 = \left(\frac{1}{D}, \dots, \frac{1}{D}\right)$, $\mathbf{a}_1$ by sampling uniformly from $\mathcal{B}_{\text{UE}}$ with out repetition, and orientation change context $\mathbf{x}_1$.
for $t \leftarrow 1, 2, \dots$ do
    #measurements and rewards:
    Measure the beams in $\mathbf{a}_q$ in order and obtain reward $\mathbf{r}_q$
    #update latent state priors:
    Obtain the updated prior for all AoA states as per (34)
    $\mathbf{p}_{q+1} \leftarrow \left(P_{1,q+1}, \dots, P_{D,q+1}\right)$
    #sample state model and select AoA:
    $\tilde{d} \leftarrow \text{mnrnd}(\mathbf{p}_{q+1}), \tilde{\phi}_q \leftarrow \phi_{\tilde{d}}$
    #select data beam for cycle $q+1$:
    Select the data beam $u^*_{q+1}$ given $\mathbf{g}$, $\tilde{\phi}_q$, and $\mathbf{x}_q$ based on the method described in Section III-C
    #select measurement subset for cycle $q+1$:
    Choose the receive beam array $\mathbf{a}_{q+1}$ as per (17) and (30)
end for
```

As shown in [7], this criterion simplies to

$$\left(b^*_{q+1}, u^*_{q+1}\right) = \underset{b \in \mathcal{B}_{\text{BS}}, u \in \mathcal{B}_{\text{UE}}}{\arg\max} \left\{ \frac{1}{N} \sum_{k=1}^{N} \log_2 \left(1 + \frac{P_{\text{tx}} d_{b,u}(t_k)}{\sigma^2}\right) \right\}, \tag{36}$$

where

$$d_{b,u}(t) = \begin{cases} \dfrac{\Omega_{b,u}\left(t, \tilde{\phi}_q\right)}{\Omega_{b,u}\left(T_{b,u}, \tilde{\phi}_q\right)} \Big[\left(1 - \rho_{b,u}\left(t, \tilde{\phi}_q\right)\right) \Omega_{b,u}\left(T_{b,u}, \tilde{\phi}_q\right) \\ \quad + \rho_{b,u}\left(t, \tilde{\phi}_q\right) g^2_{b,u}\left(T_{b,u}\right)\Big], \text{for } \rho_{b,u}\left(t, \tilde{\phi}_q\right) \geq 0, \\ \dfrac{\Omega_{b,u}\left(t, \tilde{\phi}_q\right)}{\Omega_{b,u}\left(T_{b,u}, \tilde{\phi}_q\right)} \Big[\left(1 - |\rho_{b,u}\left(t, \tilde{\phi}_q\right)|\right) \Omega_{b,u}\left(T_{b,u}, \tilde{\phi}_q\right) \\ \quad + |\rho_{b,u}\left(t, \tilde{\phi}_q\right)| (a - g_{b,u}\left(T_{b,u}\right))^2\Big], \text{else}. \end{cases} \tag{37}$$

We shall refer to the above criterion as the *AoA and predicted rate-based beam selection* (APRBS) rule. Note that $\Omega_{b,u}\left(t, \tilde{\phi}_q\right)$ and $\rho_{b,u}\left(t, \tilde{\phi}_q\right)$ are both functions of the latent AoA state $\tilde{\phi}_q$ and the user orientation $\psi(t)$. For $\rho_{b,u}\left(t, \tilde{\phi}_q\right) \geq 0$, when $g^2_{b,u}\left(T_{b,u}\right)$ or its mean $\Omega_{b,u}\left(T_{b,u}, \tilde{\phi}_q\right)$ is large, the odds that the beam pair (b, u) is selected increase. On the other hand, the reverse is true for $\rho_{b,u}\left(t, \tilde{\phi}_q\right) < 0$. As $|\rho_{b,u}\left(t, \tilde{\phi}_q\right)|$ decreases from 1 to 0, the weightage for the term that depends on $g_{b,u}\left(T_{b,u}\right)$ decreases. On the other hand, in [1], [9], [38], the data beam pair is selected as follows:

$$\left(b^*_{q+1}, u^*_{q+1}\right) = \underset{b \in \mathcal{B}_{\text{BS}}, u \in \mathcal{B}_{\text{UE}}}{\arg\max} \left\{ g^2_{b,u}\left(T_{b,u}\right) \right\}. \tag{38}$$

We shall refer to this as the *conventional power-based beam selection* (CPBS) rule.

Computational Complexity: The metric in (37) needs to be computed for B_{UE} beam pairs as per (25). For each beam pair, we need to compute $\Omega_{b,u}\left(t, \theta^{\text{rx}}_{\text{LoS}}\right)$ and $\rho_{b,u}\left(t, \theta^{\text{rx}}_{\text{LoS}}\right)$ as per (47) and (52), respectively. With this, the computation complexity equals $\mathcal{O}(B_{\text{UE}} N)$.

TABLE II
COMPARISON OF COMPLEXITY OF PROPOSED METHODS

Method	*Sampling AoA PDF*	*Beam selection for data (APRBS)*	*Selection of action*	*Prior update*
LTBS-BB	$\mathcal{O}(D)$	$\mathcal{O}(B_{\text{UE}} N)$	$\mathcal{O}(B_{\text{UE}} \log(B_{\text{UE}}))$	$\mathcal{O}(D)$
LTBS-SCM	$\mathcal{O}(D)$	$\mathcal{O}(B_{\text{UE}} N)$	$\mathcal{O}(B_{\text{UE}} \log(B_{\text{UE}}))$	$\mathcal{O}(DR)$
TV-LTBS-BB	$\mathcal{O}(D)$	$\mathcal{O}(B_{\text{UE}} N)$	$\mathcal{O}(B_{\text{UE}} \log(B_{\text{UE}}))$	$\mathcal{O}(wD)$
TV-LTBS-SCM	$\mathcal{O}(D)$	$\mathcal{O}(B_{\text{UE}} N)$	$\mathcal{O}(B_{\text{UE}} \log(B_{\text{UE}}))$	$\mathcal{O}(wDR)$

D. Regret Analysis and Complexity Comparisons

We now analyze the regret of LTBS-SCM and LTBS-BB. Given perfect knowledge of true latent state ϕ^*, let the optimal action in cycle n, which is obtained as per (16) and (17), be $\mathbf{a}^*_n$. Then, the reward associated with it is $\sum_{i=1}^{R} r_{n,\mathbf{a}^*_n(i)}$. The reward obtained by the agent is $\sum_{i=1}^{R} r_{n,\mathbf{a}_n(i)}$. The Bayes' regret $\text{BR}(q)$ up to cycle q is defined as

$$\text{BR}(q) = \mathbb{E}\left[\sum_{n=1}^{q} \sum_{i=1}^{R} \left(r_{n,\mathbf{a}^*_n(i)} - r_{n,\mathbf{a}_n(i)}\right)\right], \tag{39}$$

Proposition 1: The LTBS-SCM reward $r_{q,u}$, $\forall u \in \mathcal{B}_{\text{UE}}$, in (26) is sub-Gaussian with a proxy variance of $\tau^2_{\text{SCM}}/4$.

Proof: The proof is given in Appendix A. ■

From [39, Corr. 1], it follows that

$$\text{BR}(q) \leq \tau_{\text{SCM}} \sqrt{6Dq \log(q)} + 3D. \tag{40}$$

For LTBS-BB with a Bernoulli reward, the proxy variance can again be shown to be $1/4$. Hence, (40) applies to LTBS-BB as well.

Computational Complexity Comparison: Table II summarizes and compares the computational complexities of each step in the two variants. The prior update is the most compute-intensive step in LTBS-SCM as it involves evaluating the reward PDF R times for all the D possible AoAs. Its complexity is smaller for LTBS-BB because it uses the beta-Bernoulli conjugate pair for which the prior update rule has a simple algebraic form. The complexity of the CPBS rule, which chooses the beam pair with the largest SNR for data transmission, is $\mathcal{O}(B_{\text{UE}})$.

IV. Extension to Lateral Mobility

With lateral mobility, the AoA at the UE varies over time, resulting in a gradual beam misalignment in addition to the misalignment introduced by changes in the user orientation. This AoA variation necessitates updates to the latent state priors because the rewards and prior updates based on the previous AoA values no longer accurately reflect the current latent AoA state. In this section, we present variants of LTBS-BB and LTBS-SCM, namely TV-LTBS-BB and TV-LTBS-SCM. These variants differ from the originals in the prior update part, the other parts remain the same.

To counter the change in the AoA, we consider the observations in the past w cycles only. The smaller the value of w, the more agile is the UE to changes in the AoA as it discards older measurements. However, a small w also makes the decisions of the agent more sensitive to fading and noise.

- *TV-LTBS-BB*: Let $S_d(q)$ be the sum of rewards in cycle q for AoA direction d:
$$S_d(q) = \sum_{i=1}^{R} r_{q,\mathbf{a}_q(i)} 1_{\left\{d \in \mathcal{Q}_{q,\mathbf{a}_q(i)}\right\}}, \ \forall d \in \mathcal{D}. \tag{41}$$
In cycle q, we set the beta PDF parameters (α_d, β_d), $\forall d \in \mathcal{D}$, based on the sum of rewards in the last w cycles as follows:
$$\alpha_d = c_1 \sum_{j=q-w+1}^{q} S_d(j), \tag{42}$$
$$\beta_d = c_2 \left(wR - \sum_{j=q-w+1}^{q} S_d(j) \right), \tag{43}$$
where $c_1 > 0$ and $c_2 > 0$ are meta-parameters that control the emphasis given to the rewards. For a direction d that is not covered by any measured beam in the past w cycles, we set $\alpha_d = 1$ and $\beta_d = c_2$ to ensure that it is included in the sampling process. Its pseudo-code is similar to that in Algorithm 1 and is not shown.
- *TV-LTBS-SCM*: Similar to TV-LTBS-BB, only the rewards from cycles $q-w+1, \ldots, q$ are considered to update the prior. We, therefore, initialize the multinomial prior for cycle $q-w+1$ to $\mathbf{p}_{q-w+1} = \left(\frac{1}{D}, \ldots, \frac{1}{D}\right)$ Then, we update the prior w times as follows:
$$P_{d,k+1} = \frac{f\left(\mathbf{r}_k|\phi_d\right) P_{d,k}}{\sum_{d' \in \mathcal{D}} f\left(\mathbf{r}_k|\phi_{d'}\right) P_{d',k}}, k \in \{q-w+1, \ldots, q\}. \tag{44}$$
Now, no meta-parameters (except w) are required. The pseudo-code is similar to Algorithm 2 and is not shown.

The computational complexity of the prior updates for these variants is w times higher than that of the LTBS variants, as the update is performed w times in each cycle.

Note: The problem that we study comes under the general class of non-stationary latent bandits with context. In [40], the regret for such problems is shown to be $\mathrm{BR}\,(q) = \mathcal{O}\left(q^{\frac{2}{3}}\sqrt{DK\log(q)}\right)$, where K is the average number of latent state changes over q cycles. For a constant AoA change rate, it can be shown that $K \propto Dn$. Substituting this in the above expression, we get $\mathrm{BR}\,(q) = \mathcal{O}\left(q^{\frac{7}{6}}\right)$. Therefore, even this bound in the literature is not sub-linear in q. However, we shall see empirically that the proposed approaches converge and do so rapidly.

V. Numerical Results and Performance Benchmarking

We present the results for $B_{\text{BS}} = 18$, $B_{\text{UE}} = 18$, $N_{\text{tx}} = N_{\text{rx}} = 40$, $d^{\text{rx}} = d^{\text{tx}} = 0.25\lambda$, $v = 3.85$ kmph, and a carrier frequency of 28 GHz. Let $\eta = G_{\max}^{\text{tx}} G_{\max}^{\text{rx}} P_{\text{tx}} \Lambda / \sigma^2$ denote the peak SNR when the transmit and receive beams are aligned, where $G_{\max}^{\text{tx}} = 16$ dB and $G_{\max}^{\text{rx}} = 16$ dB are the peak transmit and receive beam gains, respectively. For SCM, the parameters are $K = 3$, $\xi_{\text{AoD}} = 10.2°$, $\xi_{\text{AoA}} = 15.5°$, $C = 4$, and $L = 20$ [34]. The beam measurement and data transmission parameters are $T_{\text{s}} = 0.125$ ms and $T_{\text{p}} = 5.14 T_{\text{s}}$. Thus, for $R = 6$ and $T_{\text{m}} = 20$ ms, the cycle duration is $T_{\text{meas}} = 120$ ms. Unless otherwise stated, the threshold τ_{BB} for LTBS-BB is set to -12 dB. It is the SNR at which the BS-UE control channels can be decoded reliably. The threshold cap τ_{SCM} in LTBS-SCM is set to 0 dB.[3] We set $D = 360$. The simulations are run over 30 SCM channel traces. Each trace is of duration 25.2 s and, thus, contains 210 cycles.

Benchmarking: We benchmark the proposed methods with the following methods:

- *Round-robin Beam Measurement (RR) [1], [9]:* The received beams are measured in a pre-defined cyclical pattern. In a cycle, let the UE receive the first pilot burst on beam k. It, thus, measures the beam pairs $(1,k), (2,k), \ldots, (B_{\text{BS}}, k)$. It receives the second pilot burst on beam $k+1$ and so on until beam $k+R-2$.[4] Lastly, it receives the R^{th} pilot burst using serving receive beam u_q^*. In the next cycle, the UE receives the pilot bursts with beams $k+R-1, k+R, \ldots$.
- *Random Beam Measurement (RND):* The first $R-1$ beams are selected randomly from $\mathcal{B}_{\text{BS}} \setminus \{u_q^*\}$ in each cycle. As above, the R^{th} pilot burst is received with serving beam u_q^*.
- *UCB* [13], [14]: The receive beams are sorted in the decreasing order of their UCB values. The $R-1$ receive beams are selected among them and sounded in the same order. The R^{th} pilot burst is received with the serving beam.
- *ϵ-Greedy* [13]: The $R-1$ receive beams with the largest received beam power values are first short-listed from among all receive beams. With probability $1-\epsilon$, each receive beam is included in the subset of beams to be measured. Else, with probability ϵ, one among the beams that are yet to be selected is included in the subset. The R^{th} pilot burst is received with the serving beam. In order to ensure as fair a comparison as possible, we have numerically fine-tuned ϵ to maximize the average data rate. The optimal value of ϵ is close to 0.15 for a wide range of system parameters.
- *Thompson Beam Sampling (TBS) [15]*: Thompson sampling with beta-Bernoulli updates in every cycle is used to select the receive beams. One sample is generated for each receive beam from its prior distribution. The $R-1$ beams from $\mathcal{B}_{\text{BS}} \setminus \{u_q^*\}$ are selected and sounded in the same order. The R^{th} pilot burst is received with the serving beam. While this approach also uses Thompson sampling, it does not model the AoA as a latent state.

In all the above methods, at the end of cycle q, the beam pair with the largest measured channel power is selected as serving beam pair as per the CPBS rule in (38). Note that these measurements are obtained over multiple cycles when $R < B_{\text{UE}}$. The training overhead of all the methods is the same because the same measurement model, which is defined in Section II-A1, is used by all the methods.

We also compare with the following genie-aided method, which provides an upper bound on the average data rate achievable by any practical, causal method. The beam pair that maximizes the data rate is selected assuming that all beam

[3] We have found in our simulations that, as τ_{SCM} increases, the rate increases and saturates after $\tau_{\text{SCM}} \geq -10$ dB. The role of τ_{SCM} is primarily technical in nature as it enables us to apply known techniques to prove convergence.

[4] In case $u_q^* \in \{k, k+1, \ldots, k+R-2\}$, then the UE moves to the next receive beam and skips receiving using u_q^*, which it will receive within the last pilot burst in the cycle.

pair gains are perfectly known at all times at both BS and UE. Thus,

$$(b^*,u^*) = \underset{b\in\mathcal{B}_{\text{BS}},u\in\mathcal{B}_{\text{UE}}}{\arg\max} \left\{\left(1-\frac{T_{\text{p}}}{T_{\text{m}}}\right) \times \sum_{k=1}^{N}\log_2\left(1+\frac{P_{\text{tx}}g_{b,u}^2\left(t_k\right)}{\sigma^2}\right)\right\}. \tag{45}$$

First, in Section V-A, we present results with only user orientation changes. In Section V-B, we present results with user orientation changes and lateral mobility. We study the impact of several system parameters such as SNR, R, and orientation change rate $\psi'\left(t\right)$, which has units of $^\circ/s$.

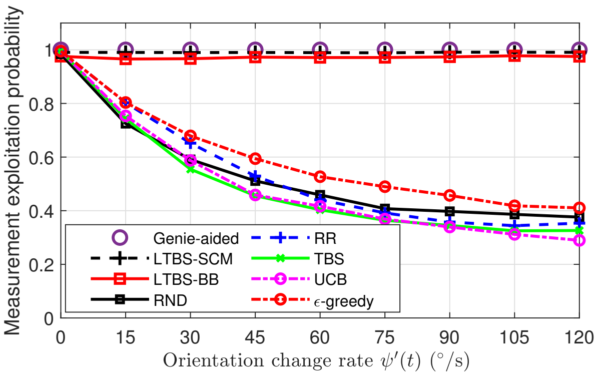

Fig. 6. Measurement exploitation probability as a function of the UE orientation change rate $\psi'\left(t\right)$ for LTBS and conventional methods ($R = 6$ and $\eta = 20$ dB).

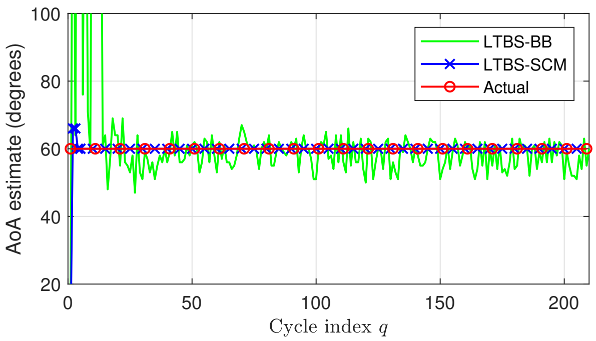

Fig. 7. Zoomed-in view of the AoA estimate as a function of cycle q ($\psi'\left(t\right) = 60^\circ$/s and $R = 6$).

A. With User Orientation Changes

Fig. 6 plots the measurement exploitation probability, which is the probability that the best beam is among the R shortlisted beams, as a function of $\psi'\left(t\right)$ for all the methods. The higher this probability, the better is the ability of the method to shortlist the receive beam that is likely to be the best one for data. This probability is always 100% for the genie-aided method. The probability for LTBS-SCM remains above 99% for all $\psi'\left(t\right)$, while that for LTBS-BB remains above 95%. Furthermore, the probabilities of both these methods are insensitive $\psi'\left(t\right)$. This is because the methods track the AoA as a latent state and choose the beams to measure and select based on it. On the other hand, the measurement exploitation probability significantly decreases for RR and RND as $\psi'\left(t\right)$ increases. Even UCB, TBS, and ϵ-greedy perform similar to RR and RND, despite being learning-based, because they do not track the AoA. For example, at $\psi'\left(t\right) = 60^\circ$/s, the measurement exploitation probabilities of RND, RR, TBS, UCB, and ϵ-greedy are 46%, 44%, 40%, 42%, and 53%, respectively. These methods are overwhelmed by the rapid changes in the beam pair gains.

Fig. 7 plots the AoA estimate $\tilde{\phi}_q$ of LTBS-BB and LTBS-SCM as a function of the cycle index q at $\psi'\left(t\right) = 60^\circ$/s. For LTBS-SCM, we observe that $\tilde{\phi}_q$ is very close to the actual AoA after just four cycles, while it takes 20 cycles for LTBS-BB. When averaged over the remaining duration, the mean error is -0.03° and error standard deviation is 0.42°, which is much smaller than the mean beamwidth of 20°. For LTBS-BB, over the same duration, the mean error and standard deviation of the error are -3.30° and 28.20°, respectively. The standard deviation is larger because of its use of the beta prior, which is not matched to the mmWave channel statistics. The less accurate selection of receive beams for measurement and weaker beam measurements leads to a larger standard deviation. LTBS-BB also requires more cycles to converge than LTBS-SCM. However, even the simpler LTBS-BB method, which has a larger standard deviation, does not fail as it achieves a high measurement exploitation probability.

Fig. 8 plots the average data rates of all the methods as a function of $\psi'\left(t\right)$, which is increased from 0°/s to 120°/s. We note that 120°/s is a high rate of change of UE orientation since it implies that a UE would complete a full rotation in just 3 s. The average data rate of LTBS-SCM is 99% of that of the genie-aided method for all $\psi'\left(t\right)$. LTBS-SCM has the highest data rate among all methods for all $\psi'\left(t\right)$. Despite using a simplistic beta prior, the LTBS-BB method has a 46%, 49%, 60%, 76%, and 51% higher data rate than RND, RR, TBS, UCB, and ϵ-greedy, respectively, at $\psi'\left(t\right) = 60^\circ$/s. As $\psi'\left(t\right)$ increases, there is a pronounced reduction in the average data rates of RR, RND, UCB, and TBS, unlike the genie-aided method and LTBS-SCM. For example, when $\psi'\left(t\right)$ increases from 30°/s to 90°/s, the average data rates of RND, RR, TBS,

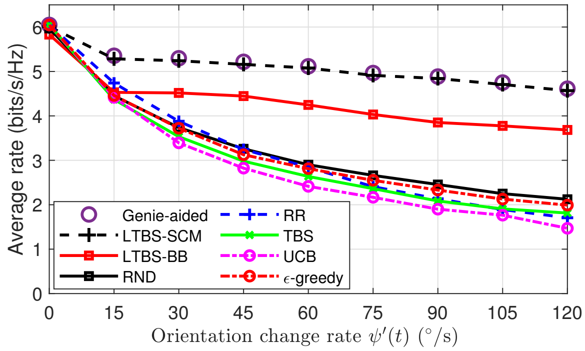

Fig. 8. Average data rate as a function of the UE orientation change rate $\psi'\left(t\right)$ ($R = 6$ and $\eta = 20$ dB).

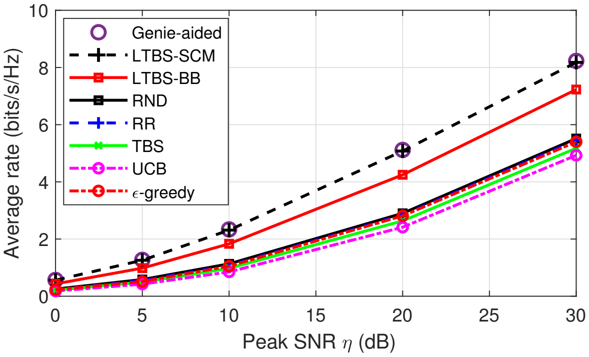

Fig. 9. Average data rate as a function of the peak SNR η ($R = 6$ and $\psi'(t) = 60°$/s).

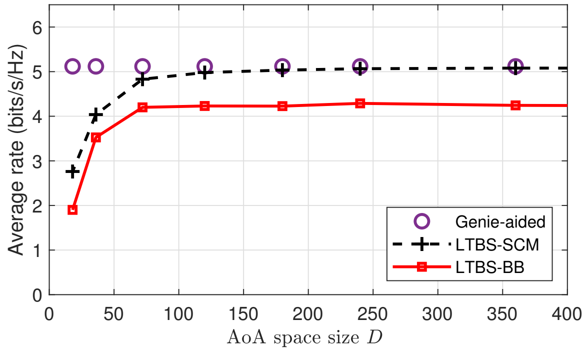

Fig. 10. Average data rates of the LTBS-BB, LTBS-SCM, and genie-aided methods as a function of D ($\psi'(t) = 60°$/s and $R = 6$).

UCB, ϵ-greedy LTBS-BB, LTBS-SCM, and the genie-aided method decrease by 34%, 45%, 41%, 44%, 37%, 15%, 7.7% and 7.7%, respectively. Thus, the LTBS methods are more robust to user orientation changes.

Fig. 9 compares the average data rates of all the methods as a function of the peak SNR η. The average data rate increases as η increases. LTBS-SCM is consistently close to the genie-aided method for all η and, thus, is near-optimal. Despite being a learning-based method, UCB has the lowest average data rate among all the methods. RR, RND, TBS, and ϵ-greedy are only marginally better than UCB, with RND being the best among the four. When η increases from 10 dB to 30 dB, LTBS-SCM achieves 114% and 50% higher rates, respectively, than RR. This is because of two reasons. First, as we saw, the odds that these methods select the optimal beam for data reception are lower. Second, LTBS-BB and LTBS-SCM use the prediction-based APRBS rule while the other methods use the CPBS rule to select the data beam pair.

Fig. 10 plots the average data rates of LTBS-BB and LTBS-SCM as a function of D. Also shown for reference is the average data rate of the genie-aided method. As D increases, the average data rates of LTBS-BB and LTBS-SCM increase. This is because of the finer quantization of the latent AoA space, which improves the accuracy of the AoA estimate and the subset of receive beams selected to measure in the next measurement cycle. The average data rate of LTBS-SCM saturates to 98.4% of that of the genie-aided method. The average data rates of both methods saturate for $D \geq 180$. They are at least 98% of the maximum value even for $D = 120$. The saturation occurs because any further increase in D does not improve the method's AoA estimation accuracy or the selection of the receive beam subset for measurement.

Fig. 11 plots the mean regret, averaged over all traces, as a function of the cycle index q for all the learning methods. For ease of comparison, the reward is determined as per (26) and the mean regret is computed as per (39) for all the methods. The mean regret of a method captures the accumulated power difference between the measurements obtained using a genie-aided approach, which utilizes the true latent state, and those obtained with the method itself. The mean regrets for LTBS-BB and LTBS-SCM decrease monotonically as q increases. This demonstrates their ability to learn the AoA. On the other hand, the mean regrets for TBS, UCB, and ϵ-greedy are larger and reach a large floor after 60 cycles.

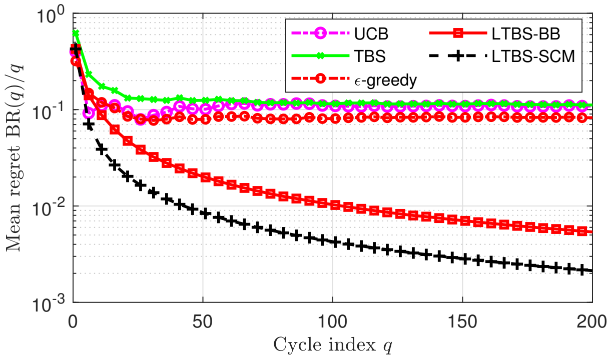

Fig. 11. Mean regret as a function of the cycle index q ($\psi'(t) = 60°$/s and $R = 6$).

B. With User Orientation Changes and Lateral Mobility

With lateral mobility, the AoA changes with time, resulting in a time-varying true latent state ϕ^*. Fig. 12 plots the estimated AoA as a function of the cycle index q for LTBS-BB and LTBS-SCM, which assume a fixed ϕ^*, and TV-LTBS-BB and TV-LTBS-SCM, which do not. We show results for an AoA change rate of 10°/s and $\psi'(t) = 60°$/s. We set $c_1 = c_2 = 10$ and $w = 6$ for TV-LTBS-BB and $w = 3$ for TV-LTBS-SCM. For TV-LTBS-SCM, the estimated AoA is very close to the actual AoA after just 3 cycles. When averaged over the remaining duration, the error mean is 1.82° and the error standard deviation is 1.38°. While these values are more than those without lateral mobility, they are still much lower than the beamwidth. The corresponding numbers for TV-LTBS-BB are 2.26° and 29.27°. The larger standard deviation of the AoA estimation error in TV-LTBS-BB is because the beta PDF parameters for multiple latent states are comparable. This leads to a higher probability that the best beam pair is not included in the measurement subset and selected for data transmission. The mean and standard deviation of the AoA estimation error for LTBS-BB are 12.47° and 33.26°, respectively, and for LTBS-SCM are 15.27° and 22.65°, respectively. As LTBS-BB and LTBS-SCM use rewards obtained from the 0^{th} cycle until

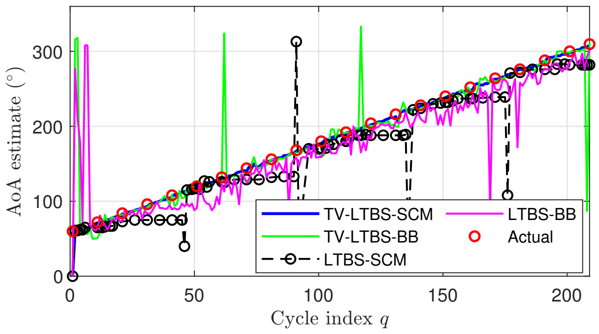

Fig. 12. Evolution of AoA and its estimate by TV-LTBS-SCM, TV-LTBS-BB, LTBS-SCM, LTBS-BB, and the actual AoA as a function of the cycle index for the latent space methods ($\psi'(t) = 60$°/s, $R = 6$, and AoA change rate of 10°/s).

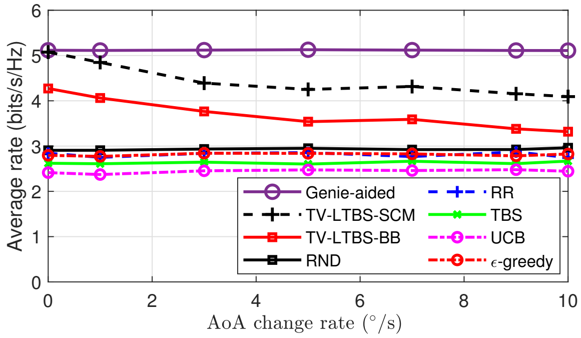

Fig. 13. Average data rate as a function of the AoA change rate for LTBS and conventional methods ($\psi'(t) = 60$°/s and $R = 6$).

the cycle under consideration, they are slow in updating their AoA estimate and end up estimating the AoA less accurately.

Fig. 13 plots the average data rate of all the methods as a function of the AoA change rate, which is increased from 0°/sec to 10°/sec. An AoA change rate of 10°/s corresponds to a minimum lateral speed of 125.6 kmph when the UE is at a distance of 200 m from the BS.[5] For example, the average data rate of TV-LTBS-SCM and TV-LTBS-BB decrease by 19.3% and 22.3%, respectively, when the AoA change rate increases from 0°/s to 10°/s. However, they still outperform the benchmark algorithms. When the AoA change rate lies in the range 0°/s to 10°/s, the average data rate of TV-LTBS-SCM is at least 39%, 49%, 53%, 68%, and 45% more than that of RND, RR, TBS, UCB, and ϵ-greedy, respectively. The same for TV-LTBS-BB are 12%, 21%, 24%, 36%, and 17%. As the AoA change rate increases, the average data rates of TV-LTBS-SCM and TV-LTBS-BB decrease because of the lag in estimating the AoA based on the measured beam gain history.

VI. Conclusions

The LTBS method employed latent Thompson sampling to estimate the AoA, which was modeled as a latent state. The sampled AoA estimate affected the subset of receive beams selected for measurements and the beam gain prediction that was used to select the beam pair for data. We saw that best beam was selected with a probability of at least 99% in LTBS-SCM and 95% in LTBS-BB. Compared to the conventional and learning-based methods proposed in the literature, both LTBS variants achieved a higher average data rate, had a lower AoA estimation error, and were less sensitive to an increase in the user orientation change rate. LTBS-SCM had a higher average data rate and a lower AoA error than LTBS-BB because of its realistic modeling of mmWave channel statistics but at the cost of a more computationally involved prior update. We also extended the LTBS methods to handle lateral mobility, which led to the AoA itself varying with time. The use of the windowing approach for updating the prior led to robust and accurate AoA tracking and a higher average data rate.

Appendix

A. Proof of Proposition 1

Since $r_{q,u}$ is a positive-valued RV that is bounded above by τ_{SCM}, we get the following inequality from the Hoeffding's lemma [41, p. 21]:

$$\mathbb{E}\left[\exp\left(\lambda(r_{q,u}-\mathbb{E}\left[r_{q,u}\right])\right)\right] \le \exp\left(\frac{\lambda^2\tau_{\text{SCM}}^2}{8}\right). \tag{46}$$

Thus, $r_{q,u}$, $\forall u \in \mathcal{B}_{\text{UE}}$, is sub-Gaussian with proxy variance of $\tau_{\text{SCM}}^2/4$.

B. MBN Model for Time-Varying SCM Channels

The MBN parameters $\Omega_{b,u}\left(t+\tau, \theta_{\text{LoS}}^{\text{rx}}\right)$, $\rho_{b,u}\left(t, \theta_{\text{LoS}}^{\text{rx}}\right)$, and m are given as follows in terms of the SCM parameters.

1) Mean Channel Power $\Omega_{b,u}\left(t, \theta_{LoS}^{rx}\right) = \mathbb{E}\left[g_{b,u}^2\left(t\right)\right]$: It is given by

$$\begin{aligned}\Omega_{b,u}\left(t, \theta_{\text{LoS}}^{\text{rx}}\right) = {} & \frac{\Lambda}{K+1}\sum_{c=1}^{C}\gamma_c \bar{G}_{u,c}^{\text{rx}}(t)\bar{G}_{b,c}^{\text{tx}}(t) \\ & + \frac{K\Lambda}{K+1}\left|Z_u^{\text{rx}}\left(\theta_{\text{LoS}}^{\text{rx}}+\psi\left(t\right)\right)\right|^2\left|Z_b^{\text{tx}}\left(\theta_{\text{LoS}}^{\text{tx}}\right)\right|^2, \end{aligned} \tag{47}$$

where the u^{th} receive beam gain $Z_u^{\text{rx}}\left(\theta\right) = \left(\mathbf{v}_u^{\text{rx}}\right)^{\dagger}\mathbf{u}_{\text{rx}}(\theta)$ is

$$\left|Z_u^{\text{rx}}\left(\theta\right)\right|^2 = \frac{1}{N_{\text{rx}}^2}\frac{\sin^2\left(N_{\text{rx}}\pi\left[\mu^{\text{rx}}\left(\theta\right)-\mu^{\text{rx}}\left(\theta_u^{\text{rx}}\right)\right]\right)}{\sin^2\left(\pi\left[\mu^{\text{rx}}\left(\theta\right)-\mu^{\text{rx}}\left(\theta_u^{\text{rx}}\right)\right]\right)}, \tag{48}$$

the b^{th} transmit beam gain $Z_b^{\text{tx}}\left(\theta\right) = \left(\mathbf{v}_b^{\text{tx}}\right)^{\dagger}\mathbf{u}_{\text{tx}}(\theta)$ is

$$\left|Z_b^{\text{tx}}\left(\theta\right)\right|^2 = \frac{1}{N_{\text{tx}}^2}\frac{\sin^2\left(N_{\text{tx}}\pi\left[\mu^{\text{tx}}\left(\theta\right)-\mu^{\text{tx}}\left(\theta_b^{\text{tx}}\right)\right]\right)}{\sin^2\left(\pi\left[\mu^{\text{tx}}\left(\theta\right)-\mu^{\text{tx}}\left(\theta_b^{\text{tx}}\right)\right]\right)}. \tag{49}$$

The receive beam gain $\bar{G}_{u,c}^{\text{rx}}(t)$ and transmit beam gain $\bar{G}_{b,c}^{\text{tx}}(t)$ for each cluster are

$$\bar{G}_{u,c}^{\text{rx}}(t) = \frac{1}{\sqrt{\pi}}\sum_{q=1}^{M} w_q\left|Z_u^{\text{rx}}\left(\sqrt{2}\sigma_{\text{AoA},c}x_q+\bar{\theta}_{\text{AoA},c}+\psi\left(t\right)\right)\right|^2, \tag{50}$$

$$\bar{G}_{b,c}^{\text{tx}}(t) = \frac{1}{\sqrt{\pi}}\sum_{q=1}^{M} w_q\left|Z_b^{\text{tx}}\left(\sqrt{2}\sigma_{\text{AoD},c}x_q+\bar{\theta}_{\text{AoD},c}\right)\right|^2, \tag{51}$$

[5] In general, if the UE is moving at a speed v m/s at an angle of θ relative to the line joining the BS and is at a distance d from the BS, then the AoA change rate is equal to $\left(\frac{v\sin(\theta)180}{\pi d}\right)$°/s.

w_q and x_q are the q^{th} Gauss-Hermite (GH) weight and abscissa, respectively, and M is the GH integration order [36, (25.4.46)].

2) Power Correlation Coefficient $\rho_{b,u}\left(t, \theta_{LoS}^{rx}\right)$ Between $g_{b,u}\left(t\right)$ and $g_{b,u}\left(t+\tau\right)$: It is defined as

$$\rho_{b,u}\left(t, \theta_{\text{LoS}}^{\text{rx}}\right) = \frac{\mathbb{E}\left[g_{b,u}^2\left(t\right) g_{b,u}^2\left(t+\tau\right)\right] - \Omega_{b,u}\left(t, \theta_{\text{LoS}}^{\text{rx}}\right)\Omega_{b,u}\left(t+\tau, \theta_{\text{LoS}}^{\text{rx}}\right)}{\sqrt{\text{Var}(g_{b,u}^2\left(t\right))\text{Var}(g_{b,u}^2\left(t+\tau\right))}}. \tag{52}$$

The term $\text{Var}(g_{b,u}^2\left(t\right)) \triangleq \mathbb{E}\left[g_{b,u}^4\left(t\right)\right] - \Omega_{b,u}^2\left(t\right)$ in the denominator is given by

$$\begin{aligned}\text{Var}(g_{b,u}^2\left(t\right)) = \frac{2\Lambda^2}{L(K+1)^2}\Bigg(&\sum_{c_1=1}^{C}\gamma_{c_1}^2 F_{u,c_1}^{\text{rx}}(t) F_{b,c_1}^{\text{tx}}(t)\\ &+2(L-1)\sum_{c_1=1}^{C}\gamma_{c_1}^2\left(\bar{G}_{u,c_1}^{\text{rx}}(t)\right)^2\left(\bar{G}_{b,c_1}^{\text{tx}}(t)\right)^2\\ &+2L\left[\sum_{c_1=1}^{C}\gamma_{c_1}\bar{G}_{u,c_1}^{\text{rx}}(t)\bar{G}_{b,c_1}^{\text{tx}}(t)\right]\sum_{c_2=1,c_2\neq c_1}^{C}\gamma_{c_2}\bar{G}_{u,c_2}^{\text{rx}}(t)\bar{G}_{b,c_2}^{\text{tx}}(t)\Bigg)\\ &+\frac{2K\Lambda^2\left|Z_u^{\text{rx}}\left(\theta_{\text{LoS}}^{\text{rx}}+\psi\left(t\right)\right)\right|^2\left|Z_b^{\text{tx}}\left(\theta_{\text{LoS}}^{\text{tx}}\right)\right|^2}{(K+1)^2}\\ &\times\sum_{c_1=1}^{C}\gamma_{c_1}\bar{G}_{u,c_1}^{\text{rx}}(t)\bar{G}_{b,c_1}^{\text{tx}}(t),\end{aligned} \tag{53}$$

where

$$F_{u,c}^{\text{rx}}(t) = \frac{1}{\sqrt{\pi}}\sum_{q=1}^{M} w_q\left|Z_u^{\text{rx}}\left(\sqrt{2}\sigma_{\text{AoA},c}x_q + \bar{\theta}_{\text{AoA},c} + \psi\left(t\right)\right)\right|^4, \tag{54}$$

$$F_{b,c}^{\text{tx}}(t) = \frac{1}{\sqrt{\pi}}\sum_{q=1}^{M} w_q\left|Z_b^{\text{tx}}\left(\sqrt{2}\sigma_{\text{AoD},c}x_q + \bar{\theta}_{\text{AoD},c}\right)\right|^4. \tag{55}$$

The term $\mathbb{E}\left[g_{b,u}^2\left(t\right) g_{b,u}^2\left(t+\tau\right)\right]$ in the numerator is derived in [7, Appendix B]. We do not show it here due to space constraints.

3) Nakagami Parameter m: Its maximum likelihood estimate is given by $m = \frac{\Omega_{b,u}^2\left(t, \theta_{\text{LoS}}^{\text{rx}}\right)}{\text{Var}(g_{b,u}^2\left(t\right))}$, where $\Omega_{b,u}\left(t, \theta_{\text{LoS}}^{\text{rx}}\right)$ is given in (47) and $\text{Var}(g_{b,u}^2\left(t\right))$ is given in (53).

C. Comparison of Marginal PDFs for $\rho \triangleq \rho_{b,u}\left(t, \theta_{LoS}^{rx}\right) < 0$

1) Comparison of PDFs in (10) and (13): Fig. 14 plots the KL divergence between the exact PDF in (13) and the Nakagami-m PDF in (10) as a function of $|\rho|$ for different values of m and Ω. The KL divergence is 0 if and only if the two PDFs are identical. The smaller the KL divergence, the tighter is the approximation. We observe that the KL divergence is less than 0.01 for $|\rho| < 0.90$ for all parameter combinations. Furthermore, the KL divergence is insensitive to Ω. Thus, the Nakagami-m PDF is a good approximation for the PDF in (13).

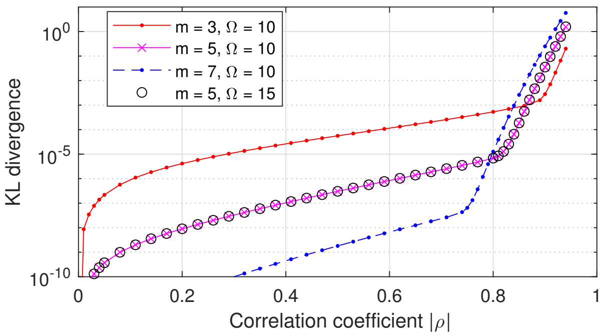

Fig. 14. KL divergence between the exact PDF in (13) and the Nakagami-m PDF as a function of $|\rho|$ for different values of m and Ω.

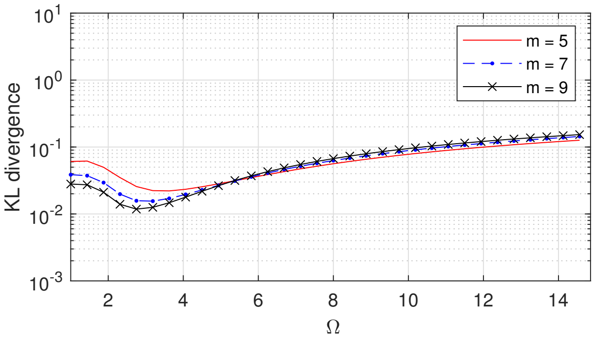

Fig. 15. KL divergence between the exact PDF in (15) and the Nakagami-m PDF for different values of m as a function of Ω.

2) Comparison of PDFs in (10) and (15): Fig. 15 plots the KL divergence as a function of Ω for different values of m.[6] It is less than 0.1 for $\Omega < 11$ for all m. Thus, the Nakagami-m PDF is a good approximation for the PDF in (15).

References

[1] M. Giordani, M. Polese, A. Roy, D. Castor, and M. Zorzi, "A tutorial on beam management for 3GPP NR at mmWave frequencies," *IEEE Commun. Surv. Tuts.*, vol. 21, no. 1, pp. 173–196, 4th Qtr. 2019.

[2] V. S. S. Ganji, T.-H. Lin, F. A. Espinal, and P. R. Kumar, "Beamsurfer: Minimalist beam management of mobile mm-wave devices," *IEEE Trans. Wireless Commun.*, vol. 21, no. 11, pp. 8935–8949, Nov. 2022.

[3] J. Rodríguez-Fernández, N. González-Prelcic, K. Venugopal, and R. W. Heath, "Frequency-domain compressive channel estimation for frequency-selective hybrid millimeter wave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 17, no. 5, pp. 2946–2960, May 2018.

[4] M. Hashemi, A. Sabharwal, C. Emre Koksal, and N. B. Shroff, "Efficient beam alignment in millimeter wave systems using contextual bandits," in *Proc. INFOCOM*, Apr. 2018, pp. 2393–2401.

[5] Z. Qi and W. Liu, "Three-dimensional millimetre-wave beam tracking based on smart phone sensor measurements and direction of arrival/time of arrival estimation for 5G networks," *IET Microw. Antennas Propag.*, vol. 12, no. 3, pp. 271–279, Jan. 2018.

[6] A. K. R. Chavva, S. Khunteta, C. Lim, Y. Lee, J. Kim, and Y. Rashid, "Sensor intelligence based beam tracking for 5G mmwave systems: A practical approach," in *Proc. Globecom*, Dec. 2019, pp. 1–6.

[7] A. K. R. Chavva and N. B. Mehta, "Millimeter-wave beam selection in time-varying channels with user orientation changes," *IEEE Trans. Wireless Commun.*, vol. 20, no. 11, pp. 6987–7000, Nov. 2021.

[8] "NR - Physical layer procedures for control," 3rd Generation Partnership Proj. (3GPP), TS 38.213, v15.6.0, 2019.

[6] Since the PDFs do not depend on $\rho_{b,u}\left(t, \theta_{\text{LoS}}^{\text{rx}}\right)$, we vary Ω and m instead.

[9] L. Wei, Q. Li, and G. Wu, "Exhaustive, iterative and hybrid initial access techniques in mmWave communications," in *Proc. WCNC*, Mar. 2017, pp. 1–6.

[10] J. Zhang, Y. Huang, Q. Shi, J. Wang, and L. Yang, "Codebook design for beam alignment in millimeter wave communication systems," *IEEE Trans. Commun.*, vol. 65, no. 11, pp. 4980–4995, Nov. 2017.

[11] M. Kim, S.-E. Hong, and J.-H. Na, "Beam selection for cell-free millimeter-wave massive MIMO systems: A matching-theoretic approach," *IEEE Wireless Commun. Lett.*, vol. 12, no. 8, pp. 1459–1463, Aug. 2023.

[12] W. Wu, D. Liu, X. Hou, and M. Liu, "Low-complexity beam training for 5G millimeter-wave massive MIMO systems," *IEEE Trans. Veh. Technol.*, vol. 69, no. 1, pp. 361–376, Jan. 2020.

[13] J. Zhang, Y. Huang, Y. Zhou, and X. You, "Beam alignment and tracking for millimeter wave communications via bandit learning," *IEEE Trans. Commun.*, vol. 68, no. 9, pp. 5519–5533, Sep. 2020.

[14] R. Gupta, K. Lakshmanan, and A. K. Sah, "Beam alignment for mmWave using non-stationary bandits," *IEEE Commun. Lett.*, vol. 24, no. 11, pp. 2619–2622, Nov. 2020.

[15] Y. Wang, K. Zu, W. Zheng, and L. Zhang, "Fast mmwave beam alignment method with adaptive discounted Thompson sampling," in *Proc. WCNC*, Apr. 2024, pp. 1–6.

[16] L. Wei, Q. Li, and G. Wu, "Exhaustive, iterative and hybrid initial access techniques in mmWave communications," in *Proc. WCNC*, Mar. 2017, pp. 1–6.

[17] D. Ghosh, M. K. Hanawal, and N. Zlatanov, "UB3: Fixed budget best beam identification in mmwave massive MISO via pure exploration unimodal bandits," *IEEE Trans. Wireless Commun.*, vol. 23, no. 10, pp. 12 658–12 669, Oct. 2024.

[18] N. Blinn, J. Boerger, and M. Bloch, "mmwave beam steering with hierarchical optimal sampling for unimodal bandits," in *Proc. ICC*, Jun. 2021, pp. 1–6.

[19] I. Aykin, B. Akgun, M. Feng, and M. Krunz, "MAMBA: A multi-armed bandit framework for beam tracking in millimeter-wave systems," in *Proc. INFOCOM*, Jul. 2020, pp. 1469–1478.

[20] S. Sarkar, M. Krunz, I. Aykin, and D. Manzi, "Machine learning for robust beam tracking in mobile millimeter-wave systems," in *Proc. Globecom*, Dec. 2021, pp. 1–6.

[21] G. Ghatak, "Best arm identification based beam acquisition in stationary and abruptly changing environments," *IEEE Trans. Signal Process.*, vol. 72, no. 1, pp. 670–685, Jan. 2024.

[22] A. Kumar, A. Roy, and R. Bhattacharjee, "Actively adaptive multi-armed bandit based beam tracking for mmwave mimo systems," in *Proc. WCNC*, Apr. 2024, pp. 1–6.

[23] W. Wu, N. Cheng, N. Zhang, P. Yang, W. Zhuang, and X. Shen, "Fast mmwave beam alignment via correlated bandit learning," *IEEE Trans. Wireless Commun.*, vol. 18, no. 12, pp. 5894–5908, Dec. 2019.

[24] Y. Zhang, S. Basu, S. Shakkottai, and R. W. Heath, "Mmwave codebook selection in rapidly-varying channels via multinomial Thompson sampling," in *Proc. ACM MobiHoc*, Jul. 2021, p. 151–160.

[25] P. Susarla, B. Gouda, Y. Deng, M. Juntti, O. Silvén, and A. Tölli, "Learning-based beam alignment for uplink mmwave uavs," *IEEE Trans. Wireless Commun.*, vol. 22, no. 3, pp. 1779–1793, Mar. 2023.

[26] C. Liu, L. Zhao, M. Li, and L. Yang, "Adaptive beam search for initial beam alignment in millimetre-wave communications," *IEEE Trans. Veh. Technol.*, vol. 71, no. 6, pp. 6801–6806, Jun. 2022.

[27] S. H. A. Shah and S. Rangan, "LSTM-aided selective beam tracking in multi-cell scenario for mmwave wireless systems," *IEEE Trans. Wireless Commun.*, vol. 23, no. 2, pp. 890–907, Feb. 2024.

[28] M. B. Booth, V. Suresh, N. Michelusi, and D. J. Love, "Multi-armed bandit beam alignment and tracking for mobile millimeter wave communications," *IEEE Commun. Lett.*, vol. 23, no. 7, pp. 1244–1248, Jul. 2019.

[29] "Study on channel model for freq. from 0.5 to 100 GHz," 3rd Gen. Partnership Proj. (3GPP), TR 38.901, v14.2.2, 2017.

[30] A. Slivkins, "Introduction to multi-armed bandits," *Foundations and Trends in Machine Learning*, vol. 12, no. 1-2, pp. 1–286, 2019.

[31] A. K. I. O. Daniel J. Russo, Benjamin Van Roy and Z. Wen, *A Tutorial on Thompson Sampling*. NOW: Foundations and Trends in Machine Learning, 2018.

[32] O. Chapelle and L. Li, "An empirical evaluation of thompson sampling," in *Adv. in Neural Inf. Process. Syst.*, J. Shawe-Taylor, R. Zemel, P. Bartlett, F. Pereira, and K. Weinberger, Eds., vol. 24. Curran Associates, Inc., 2011.

[33] "Technical specification: High precision 6-axis MEMS motion tracking device ICM-42688-P," InvenSense, 2023.

[34] M. R. Akdeniz, Y. Liu, M. K. Samimi, S. Sun, S. Rangan, T. S. Rappaport, and E. Erkip, "Millimeter wave channel modeling and cellular capacity evaluation," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1164–1179, Jun. 2014.

[35] I. A. Hemadeh, K. Satyanarayana, M. El-Hajjar, and L. Hanzo, "Millimeter-wave communications: Physical channel models, design considerations, antenna constructions, and link-budget," *IEEE Commun. Surv. Tuts.*, vol. 20, no. 2, pp. 870–913, 2nd Qtr. 2018.

[36] M. Abramowitz and I. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, 9th ed. Dover, 1972.

[37] L. S. Gradshteyn and L. M. Ryzhik, *Tables of Integrals, Series and Products*, 7th ed. Academic Press, 2007.

[38] C. Jeong, J. Park, and H. Yu, "Random access in millimeter-wave beamforming cellular networks: Issues and approaches," *IEEE Commun. Mag.*, vol. 53, no. 1, pp. 180–185, Jan. 2015.

[39] J. Hong, B. Kveton, M. Zaheer, Y. Chow, A. Ahmed, and C. Boutilier, "Latent bandits revisited," in *Proc. NeurIPS*, vol. 33, Dec. 2020, pp. 13 423–13 433.

[40] J. Hong, B. Kveton, M. Zaheer, Y. Chow, A. Ahmed, M. Ghavamzadeh, and C. Boutilier, "Non-stationary latent bandits," *CoRR*, vol. abs/2012.00386, Dec. 2020.

[41] P. Massart, *Concentration Inequalities and Model Selection*, 1^{st} ed. Springer, 2003.

[42] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2^{nd} ed. Wiley Series in Telecommunications, 2005.