

# Beam Drift in Millimeter Wave Links: Beamwidth Tradeoffs and Learning Based Optimization

Jianjun Zhang, *Member, IEEE*, and Christos Masouros, *Senior Member, IEEE*

**Abstract**—Millimeter wave (mmwave) communications, envisaged for the next generation wireless networks, rely on large antenna arrays and very narrow, high-gain beams. This poses significant challenges to beam alignment between transmitter and receiver, which has attracted considerable research attention. Even when alignment is achieved, the link is subject to beam drift (BD). BD, caused by non-ideal features inherent in practical beams and rapidly changing environments, is referred to as the phenomenon that the center of main-lobe of the used beam deviates from the real dominant channel direction, which further deteriorates the system’s performance. To mitigate the BD effect, in this paper we first theoretically analyze the BD effect on the performance of outage probability as well as effective achievable rate, which takes practical factors (e.g., the rate of change of the environment, beam width, transmit power) into account. Then, different from conventional practice, we propose a novel design philosophy where multi-resolution beams with varying beam widths are used for data transmission while narrow beams are employed for beam training. Finally, we design an efficient learning based algorithm which can adaptively choose an appropriate beam width according to the environment. Simulation results demonstrate the effectiveness and superiority of our proposals.

**Index Terms**—Beam drift, Bayesian contextual bandit, performance analysis, beam training, millimeter wave communication.

## I. INTRODUCTION

**T**HANKS to abundant bandwidth resources of mmwave, mmwave communications have recently attracted considerable attention [1]. However, the propagation features of mmwave signals are different from the conventional microwave counterparts, which makes it difficult to reap the benefits of mmwave communications. In particular, compared to microwave channels, the path-loss of mmwave channels is much larger. Fortunately, the wave-length of mmwave signals is small, and thus it is convenient to pack a large number of antennas into a small space, which provides large array gains via beamforming to combat the large path-loss. However, the large number of antennas poses great difficulties in obtaining channel state information (CSI) via beam training, especially in dynamic environments.

In view of the sparsity and large dimension of mmwave channels, an effective approach to obtain CSI is beam training [2]–[4]. In beam training, candidate beams at the transmitter and/or receiver are directly trained via exhaustive or adaptive

search to select the ones that optimize some performance metric, e.g., the strength of received signals or signal-to-noise ratio (SNR) [3]–[5]. In particular, as a non-adaptive approach, the hierarchical search based beam training approach has been studied deeply and a variety of algorithms have been proposed to design hierarchical codebooks [2]–[4], [6]–[8]. Nevertheless, the training overhead of the hierarchical search based beam training algorithms is still very high in systems with large-scale antenna arrays. Furthermore, beam tracking techniques can be used to track beams in some cases (e.g., when channel correlation is available), so as to further reduce the training overhead [9]–[12]. Although beam tracking techniques can reduce the training overhead, for most practical scenarios the assumptions of channel modeling required by channel tracking are often too stringent to meet [12].

To reduce the training overhead, the core approach is to reduce the beam search space for the training process via extracting and exploiting the information from the training history, which leads to machine learning (ML) based beam training algorithms [12]–[21]. According to the underlying ML approaches or principles, these beam training solutions roughly fall into two categories. The first category is designed based on supervised learning and occupies most of the existing ML based methods [12]–[17]. The key of the supervised learning based beam training solutions is to prepare for a sufficiently large database of training samples. As for the second category, it is designed based on reinforcement learning, or more generally, Markov decision process [18]–[21]. An important benefit of the second category is that the burden of collecting training samples can be relieved to a great extent.

Note that in previous studies and also in practical system designs, it is almost always assumed that channels (also channel directions) remain unchanged within any given time-slot, and most of the subsequent system designs are based on this assumption. However, this assumption may be problematic, since it ignores beam drift (BD) effect. BD is referred to as the phenomenon that angle of departure (AoD) or angle of arrival (AoA) deviates from the beam center angle aligned at the beginning stage of each time-slot, which leads to a low beamforming (or array) gain. The causes of BD are two-fold. On the one hand, AoD or AoA, which depends only on exterior physical environments (e.g., transmitter and/or receiver and their motions), almost always changes continuously. On the other hand, the beams used for data transmission (e.g., the DFT beams constructed by sampling the beam space) are non-ideal (e.g., large transition band and ripples), which results in a low beamforming gain even if the AoD or AoA deviates from only a bit the beam center angle. In fact, as long as the AoD

The work was supported by the Engineering and Physical Sciences Research Council, UK under project EP/S028455/1. (Corresponding author: Jianjun Zhang.)

J. Zhang and C. Masouros are with the Department of Electronic & Electrical Engineering, University College London, London WC1E7JE, U.K. (E-mail: {jianjun.zhang,c.masouros}@ucl.ac.uk).

or AoA changes and thus invokes beam switching, the beam center angle aligned at the beginning stage of each time-slot is bound to deviate from the AoD or AoA, which has already led to a low beamforming gain.

The BD effect poses great impacts on system performance. In particular, it affects the performance of effective achievable rate (EAR), i.e., the training overhead is also taken into account. On one hand, it has been widely accepted that narrow (and even the most narrow) beams should be used for data transmission, so as to achieve the highest beamforming gain. Nevertheless, from a beam training point of view, narrow beams increase the size of beam search space and the resulting training overheads. Moreover, the BD effect also has a negative impact on beamforming gain. This is particularly pronounced once the environment changes rapidly. In fact, as we will point out in this paper, the BD effect implies a tradeoff between the beamforming gain and effective time of data transmission. This is particularly challenging, as the BD effect is almost inevitable in mobile scenarios. Although narrowing the length of each time-slot can mitigate the BD effect, it may also reduce the EAR performance since beam training is almost essential and occupies a considerable part of each time-slot.

To mitigate the BD effect and ensure a high beamforming gain all the time, in this paper we propose a novel design philosophy for mmwave beam training and data transmission, along with an efficient algorithm to adapt the beam width. Specifically, we first theoretically analyze the performances of outage probability (OP) and EAR, by incorporating practical factors. Then, to mitigate the BD effect and improve system performance, a novel design philosophy for mmwave communication is proposed, where multi-resolution and narrow beams are used for data transmission and beam training, respectively. Finally, to enable flexible and fast beam width switching, we further design an efficient learning based algorithm. The main contributions are summarized as follows:

- We reveal the BD effect with the aim of improving the EAR performance. In particular, we derive the analytical expressions for OP and EAR, which reveal useful insights via explicitly characterizing the relationship between OP (or EAR) and practical factors, especially, the beam width and the rate of change of the environment.
- We propose a novel design philosophy for data transmission and beam training in mmwave communications. Specifically, multi-resolution beams are advocated for data transmission so as to mitigate the BD effect, while narrow beams are employed to perform beam training which accommodate the beam change.
- To mitigate the BD effect, we further design a learning based algorithm to achieve the goal of adapting the beam widths for data transmission. The designed algorithm can sense the rate of change of the environment quickly and adaptively choose the optimal beam width from a multi-resolution codebook.
- Comprehensive simulation results are provided to demonstrate the effectiveness and superiority of the proposed algorithm. It is shown that the proposed algorithm can effectively mitigate the BD effect by intelligently choosing the optimal beam according to the environment, and

achieve a good EAR performance.

The remainder of this paper is organized as follows. System model of mmwave communication is described in Section II. In Section III, analytical expressions are derived for both OP and EAR performances, and useful insights are also revealed. A novel design philosophy for mmwave data transmission and beam training is proposed in Section IV. To mitigate the BD effect, a learning based algorithm is also proposed. Simulation results and conclusions are given in Section V and Section VI, respectively. To improve readability, the proofs of all theorems are deferred to appendices.

Notations: Bold uppercase  $\mathbf{A}$  and bold lowercase  $\mathbf{a}$  denote matrices and column vectors, respectively. Without particular specification, non-bold letters  $A, a$  denote scalars. Calligraphic letters  $\mathcal{A}$  stand for sets.  $\mathbb{E}(\cdot)$  and  $(\cdot)^H$  denote the mathematical expectation and Hermitian operators, respectively.  $\mathbb{I}\{\cdot\}$  and  $\text{card}(\mathcal{A})$  represent the indicator function and the cardinality of  $\mathcal{A}$ , respectively.  $(\cdot)^*$  represents an optimal quantity, e.g., an optimal solution of an optimization problem.  $\mathcal{CN}(\mathbf{m}, \mathbf{R})$  stands for a complex Gaussian random vector with mean  $\mathbf{m}$  and covariance matrix  $\mathbf{R}$ . For an integer  $n > 0$ ,  $[n]$  is the set containing all integers from 1 to  $n$ , i.e.,  $[n] = \{1, 2, 3, \dots, n\}$ .  $\mathbf{1}$  and  $\mathbf{I}$  denote an all-one vector and an identity matrix of appropriate dimension, respectively.

## II. SYSTEM MODEL

Consider a mmwave point-to-point communication system, which consists of one base station (BS) equipped with  $N$  transmit antennas and a single-antenna user (UE), as illustrated in Fig. 1. To facilitate practical system implementation, a codebook based beamforming is considered in this paper [5]. Temporarily, we assume that each transmitting beam is chosen from a predefined codebook  $\mathcal{C} = \{\mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_M\}$  of size  $M$  (e.g., the DFT codebook constructed by uniformly sampling beam space  $[-1, 1]$ ). Later, we will consider a multi-resolution codebook to improve system performance.

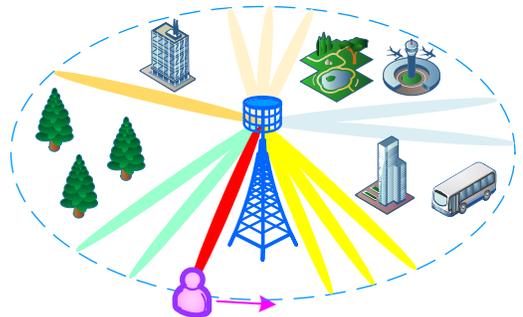


Fig. 1. An illustration of a point-to-point mmwave communication system.

Due to the sparsity of mmwave channels, an extended Saleh-Valenuela geometric channel model is adopted in this paper. The channel vector  $\mathbf{h} \in \mathbb{C}^{N \times 1}$  between the BS and the UE is given by

$$\mathbf{h} = \sqrt{N/\beta} \sum_{l=1}^L g_l \mathbf{a}(\phi_l), \quad (1)$$

where  $\mathbf{a}(\cdot)$  represents array response vector,  $\beta$  is average path-loss,  $L$  is the number of paths, and  $g_l$  is complex small-scale fading of the  $l$ -th path. In (1),  $\phi_l = \cos(\theta_l)$ , where  $\theta_l$  is physical angle of departure (AoD) of the  $l$ -th path. Without loss of generality, uniform linear array (ULA) is considered in this paper for simplicity. However, the proposed algorithms can be extended to other types of antenna array geometries, e.g., uniform planar array.

With the assumption that beam  $i$  (i.e.,  $\mathbf{f}_i \in \mathbb{C}^{N \times 1}$ ) is chosen by the BS, the signal received at the UE is given by

$$y_i = \sqrt{P} \mathbf{h}^H \mathbf{f}_i s + n_i, \quad (2)$$

where  $P$  denotes transmit power,  $s \in \mathbb{C}$  denotes pilot symbol (e.g.,  $s = 1$ ), and  $n_i \sim \mathcal{CN}(0, 1)$  denotes random noise variable. The SNR is calculated as

$$\text{SNR} = P |\mathbf{h}^H \mathbf{f}_i|^2. \quad (3)$$

Let  $T_B$  represent the duration of beam training within a time-slot and  $T_S$  be the duration of any time-slot. The metric EAR for a time-varying channel environment is defined as

$$R_E = T_S^{-1} \int_{T_B}^{T_S} \log(1 + \text{SNR}(t)) dt. \quad (4)$$

Note that  $T_S^{-1}$  in (4) is introduced to keep consistent with the definition in the literatures [12].<sup>1</sup> The definition in (4) is extended from the conventional case (i.e.,  $\text{SNR}(t)$  is a constant) to more general cases (e.g.,  $\text{SNR}(t)$  is a function), by leveraging the technique of taking the limit. Since no extra or special assumption is involved, it can apply to various complicated cases.

Apparently, to achieve a high throughput, one the one hand,  $\text{SNR}(t)$  should be as large as possible in the whole phase of data transmission, and on the other hand,  $T_B$  should be as small as possible, so as to reserve more effective time for data transmission. Note that as a time-varying quantity occurring in the phase of data transmission,  $\text{SNR}(t)$  is significantly affected by the BD effect. Meanwhile,  $T_B$  is closely related to beam training. Therefore, to improve the EAR performance, both data transmission and beam training should be carefully designed, which is the focus of this paper.

### III. PERFORMANCE ANALYSIS WITH BEAM DRIFT

In this section, we construct a mathematical model for BD via Gaussian processes (GPs) and mathematically analyze its influence on the system EAR performance.

#### A. Beam Drift and Mathematical Modeling

To facilitate study, it is often assumed that the AoDs  $\{\phi_l\}$  keep fixed within each time-slot, while vary across different time-slots only, as shown in Fig. 2-(a).<sup>2</sup> However, the AoDs,

<sup>1</sup>In fact, without considering the BD effect, the definition in (4) is simplified as  $R_E = (1 - T_B/T_S) \log(1 + \text{SNR})$ , which was adopted in [12]. Note that  $(1 - T_B/T_S)$  is incorporated to take the training overhead into account.

<sup>2</sup>In practice,  $\theta$  (in the angular domain) and  $\phi = \cos(\theta)$  (in the beam domain) are both referred to as AoD. The unit of  $\theta$  is the radian, and thus the range is  $[0, \pi]$  (or  $[-\pi, \pi]$ ), while the range of  $\phi$  is  $[-1, 1]$ . Similar to [4], AoD/AoA and beam pattern are characterized in the beam domain in this paper (e.g., to characterize beam width in (9)).

in fact, often continuously change from one time-slot to the next one, because the UE moves continuously. As a result, although the chosen beam is optimal at the beginning of each time-slot, it gradually becomes sub-optimal and even interrupts the communication link (i.e., the beamforming or array gain is less than a predefined threshold), as shown in Fig. 2-(b). In this paper, this phenomenon is referred to as beam drift.

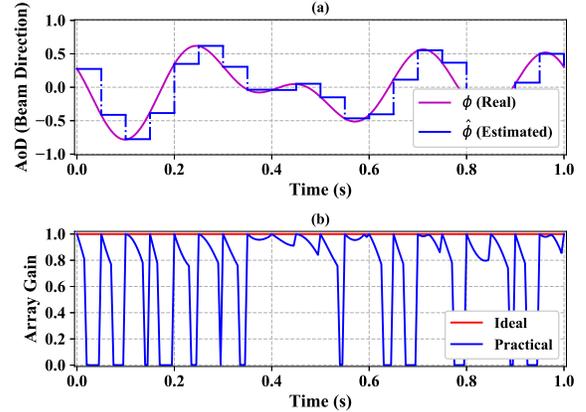


Fig. 2. (a) Beam drift phenomenon - the real AoD  $\phi(t)$  continuously changes with time, while the estimated AoD  $\hat{\phi}(t)$  keeps fixed within each time-slot; (b) Beam drift may cause a dramatic change of normalized beamforming (or array) gain, calculated as  $|\mathbf{a}^H(\phi)\mathbf{a}(\phi)|$ .  $\phi(t)$  is sampled as per a GP with kernel  $k_{\text{SEE}}(t, t') = \exp(-|t - t'|^2)$  (See the next page for more details).

It is not difficult to understand that BD has a significant impact on system performance. Typically, a larger BD leads to a larger performance loss. Because of its significance, an in-depth study of BD, in particular, to analyze its influence on performance metrics, is both required and urgent. To this end, an appropriate mathematical model has to be chosen to better characterize BD. To accommodate randomness and analytical tractability, stationary GPs are considered in this paper [22]. Moreover, since a GP is a non-parametric (stochastic) model, it provides sufficient universality and flexibility.

A GP  $\{f(t)\}$  is a collection of random variables, any finite number of which are jointly Gaussian [22]. Specifically, for any finite number of points  $t_1, t_2, \dots, t_n$ , the joint probability density function  $p(f(t_1), f(t_2), \dots, f(t_n))$  is Gaussian. A GP is completely specified by its mean function  $m(t) = \mathbb{E}[f(t)]$  and covariance function  $k(t, t')$  (also referred to as kernel in the literatures). Without loss of generality, the mean function is assumed to be zero, i.e.,  $m(t) = 0$ . The covariance function or kernel is defined as

$$\begin{aligned} k(t, t') &= \mathbb{E}[(f(t) - m(t))(f(t') - m(t'))] \\ &= \mathbb{E}[f(t)f(t')]. \end{aligned} \quad (5)$$

Accordingly, the GP is denoted by  $f(t) \sim \mathcal{GP}(k(t, t'))$ . The kernel is simplified as  $k(t)$  at a diagonal point  $t = t'$ .

The kernel is crucial for a GP, because it encodes the prior about the function. The specification of the kernel implies a distribution over the function and defines similarity or nearness of the function values at different points. Generally, an arbitrary function of input pairs  $t$  and  $t'$  may not be an effective kernel. To meet different applications (or scenarios), many kinds

of kernels have been developed, e.g., Matern kernel, squared exponential (SE) kernel, periodic kernel, spectral mixture (SM) kernel. Please refer to [23] for more details. Next, we briefly introduce two kinds of kernels, which have been widely used in practice. Note that new kernels can be constructed from the existing ones via simple mathematical operations, e.g., sum or product operations. Typically, if  $\{k_i(\cdot, \cdot) | i = 1, \dots, n\}$  are kernels,  $\sum_{i=1}^n k_i(\cdot, \cdot)$  and  $\prod_{i=1}^n k_i(\cdot, \cdot)$  are also kernels.

1) *Squared Exponential Kernel*: The SE kernel takes the form

$$k_{\text{SE}}(t, t') = \sigma_f^2 \exp\left(-\frac{1}{2l^2}|t - t'|^2\right), \quad (6)$$

where  $\sigma_f^2$  and  $l$  are signal variance and length-scale, respectively [23]. In particular, parameter  $l$  characterizes the rate of change of the GP. If the GP varies rapidly, the length-scale  $l$  should be shorter. On the other hand, the degree of variation of the GP can be achieved by adjusting  $l$ .

2) *Spectral Mixture Kernel*: The SM kernel provides more flexibility and takes the form

$$k_{\text{SM}}(t, t') = \sum_{q=1}^Q a_q \frac{|\Sigma_q|^{0.5}}{(2\pi)^{0.5}} \exp\left(-\frac{1}{2}\|\Sigma_q^{0.5}(t - t')\|^2\right) \cos(2\pi\mu_q(t - t')), \quad (7)$$

where  $\{a_q\}$ ,  $\{\Sigma_q\}$  and  $\{\mu_q\}$  are mixture weights, bandwidths (inverse length-scales) and frequencies, respectively. The SM kernel is more expressive, and helps to discover interesting and important structures and/or models of a stochastic process (e.g., quasi-periodic stationary mode).

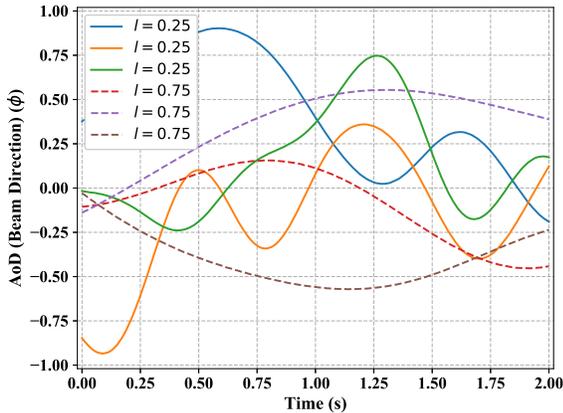


Fig. 3. Sample functions of a GP with the SE kernel -  $\sigma_f^2 = 1$ . A GP (with  $m(t) = 0$ ) is uniquely characterized by the kernel, based on which sample functions can be generated. Please refer to [22] or [23] (e.g., Appendix A in [23]) for more details.

For intuition, some sample functions of a GP with the SE kernel are shown in Fig. 3. It is observed that a smaller value of the parameter  $l$  corresponds to an environment that changes more rapidly, which confirms the effectiveness of using GPs to model AoD/AoA change. Another benefit of modeling AoD/AoA change via GPs is that it enables us to derive analytical expressions and further reveal useful insights. Note that since the SE kernel is infinitely differentiable, a GP with this kernel is smooth, as shown in Fig. 3.

## B. Performance Analysis of Beam Drift

Now, we analyze the BD effect on system performance. To simplify performance analysis and reveal meaningful insights, the number of channel paths  $L$  in (1) is assumed to be 1, i.e.,  $L = 1$ . Accordingly, the channel vector  $\mathbf{h}$  is simplified as

$$\mathbf{h} = \sqrt{N/\beta} \mathbf{g} \mathbf{a}(\phi),$$

where all subscripts are omitted for convenience. The rationale of the assumption is that for mmwave channels with the LOS link (e.g., in the scenario of fixed wireless access), the effect of NLOS links is usually marginal, compared to that of the LOS link [24]. Moreover, although the conclusions are derived based on the assumption, they still hold for more general cases.

For simplicity, we consider an ideal codebook temporarily. In Section V, practical codebooks will be considered. For an ideal codebook, each beam has an ideal beam pattern, e.g., a beam (or codeword)  $\mathbf{f}$  has the following beam pattern

$$|\mathbf{a}^H(x)\mathbf{f}| = \begin{cases} G_M & (x \in \mathcal{I}_M) \\ G_S & (x \in \mathcal{I}_S), \end{cases} \quad (8)$$

where  $G_M$  and  $G_S$  represent beamforming gains in the main-lobe  $\mathcal{I}_M$  and side-lobe  $\mathcal{I}_S$ , respectively.<sup>3</sup> In practice, the beamforming gain in the side-lobe  $\mathcal{I}_S$  is often very small, which can be approximated to 0 [25] (i.e.,  $G_S = 0$ ), so as to simplify performance analysis. Note that  $G_M$  depends on multiple factors and is not a constant. In particular, as the beam width increases,  $G_M$  decreases accordingly. To explicitly characterize this feature,  $G_M$  is approximated as [25]

$$G_M \approx \frac{1}{2b}, \quad (9)$$

where  $2b$  represents the width of main-lobe  $\mathcal{I}_M$ .

Next, we derive analytical expressions for OP and EAR, which takes the BD effect into account. First, we focus on the derivation of OP which is formally defined as

$$P_{\text{out}}(T) = \mathbb{P}(\text{SNR} < T), \quad (10)$$

where  $T$  denotes the threshold. According to (3), the SNR can be calculated as

$$\text{SNR} = \begin{cases} PN|g|^2\beta^{-1}G_M & f(t) \in \mathcal{I}_M \\ PN|g|^2\beta^{-1}G_S = 0 & f(t) \in \mathcal{I}_S, \end{cases} \quad (11)$$

where  $f(t) \in \mathcal{I}_M$  (or  $f(t) \in \mathcal{I}_S$ ) indicates that the direction of AoD/AoA at time  $t$  is within the main-lobe (or side-lobe) of beam  $\mathbf{f}$ . The OP  $P_{\text{out}}(T)$  is characterized in the following theorem.

**Theorem 1.** *Let  $2b$  denote the beam width of the used beam. If  $PN|g|^2\beta^{-1}G_M \geq T$ , the OP is upper bounded by*

$$P_{\text{out}}(T) \leq \frac{I}{2\pi} \sqrt{\frac{-2k''(0)}{k(0)}} \exp\left(-\frac{b^2}{4k(0)}\right). \quad (12)$$

<sup>3</sup>An ideal or (non-ideal) codebook consists of multiple ideal (or non-ideal) beams. The beam pattern of a non-ideal beam consists of three parts, i.e., main-lobe  $\mathcal{I}_M$ , side-lobe  $\mathcal{I}_S$  and transition band  $\mathcal{I}_T$ , such that  $\mathcal{I}_M \cup \mathcal{I}_S \cup \mathcal{I}_T = [-1, 1]$ ,  $\mathcal{I}_M \cap \mathcal{I}_S = \emptyset$ ,  $\mathcal{I}_M \cap \mathcal{I}_T = \emptyset$  and  $\mathcal{I}_S \cap \mathcal{I}_T = \emptyset$  hold and the ripples in the main-lobe and side-lobe are greater than zero. To facilitate performance analysis and derive useful insights, we mainly consider ideal beams/codebook, whose transition bands are absent and ripples are zero.

where  $I = T_S - T_B$  is the time or duration of effective data transmission within a time-slot.

*Proof:* See Appendix A. ■

For practical non-ideal beams, similar expressions can be derived as well, with some modifications. Theoretically, Theorem 1 shows that the beam width has a significant influence on system OP performance. In particular, wider beams help to achieve a lower OP (and therefore a better reliability), which coincides with our intuition.

**Remark 3.1** Theorem 1 also indicates a promising solution to keep a low OP, i.e., utilizing wide beams. In practice, the beam width can be changed flexibly (typically, by choosing an appropriate beam from a multi-resolution codebook).

In general, the more quickly the environment changes, the larger  $-k''(0)$  is [26]. Hence, Theorem 1, in fact, shows that a fast-changing environment can incur a higher OP, which is undesirable in practice. Theorem 1 also indicates that a shorter time-slot (i.e., a smaller  $T_S$ ) helps to achieve a lower OP, which coincides with our intuition. However, since beam training also occupies a considerable part of the time resource, it is not a good solution to tackle this issue by simply shortening  $T_S$ , because it will lead to a low EAR performance.

In practice, we may also be interested in EAR performance. Next, we proceed to analyze the EAR performance, which is characterized in the following theorem.

**Theorem 2.** *With the assumption of given beam width, path gain, transmit power and training overhead (which guarantees successful alignment), the expected EAR is lower bounded by*

$$\mathbb{E}(R_E) \geq \left(1 - \frac{T_B}{T_S}\right) \left(1 - \frac{1}{2\pi} \sqrt{\frac{-2k''(0)}{k(0)}} \exp\left(-\frac{b^2}{4k(0)}\right)\right) \cdot \log\left(1 + PN|g|^2\beta^{-1}G_M\right). \quad (13)$$

*Proof:* See Appendix B. ■

Note that the derived bounds in Theorem 1 and Theorem 2 are attainable (i.e., the inequalities in the two theorems are, in fact, equalities) if  $T_S$  is not large. In fact, the key to derive these bounds is to count the number of up-crossings (See the appendices for details). However, if  $T_S$  is not large, the number of up-crossings is almost always at most 1. In this case, these inequalities degenerate to the corresponding equalities.

In view of the approximation  $G_M \approx 1/(2b)$ , the expression of the expected EAR can be further written as

$$\mathbb{E}(R_E) \gtrsim \left(1 - \frac{T_B}{T_S}\right) \left(1 - \frac{1}{2\pi} \sqrt{\frac{-2k''(0)}{k(0)}} \exp\left(-\frac{b^2}{4k(0)}\right)\right) \cdot \log\left(1 + \frac{PN|g|^2}{2b\beta}\right). \quad (14)$$

It can be seen a clear tradeoff between the OP and EAR performance revealed in Theorem 1 and Theorem 2. In particular, although wide beams can reduce the OP, their beamforming gains are relatively small, which may deteriorate the EAR performance finally. Therefore, from the perspective of optimizing the EAR performance, it is important to adaptively choose a beam of optimal beam width.

An apparent method to obtain the optimal EAR performance is to maximize  $R_E$  in (13), maybe with a constraint (e.g., the OP is no greater than a preassigned value  $\varepsilon$ ). If  $k''(0)$ ,  $k(0)$ ,  $|g|$  and  $T_B$  are known or can be estimated from empirical data or can be computed analytically, it is sufficient to directly maximize  $\mathbb{E}(R_E)$  by solving an optimization problem. However, practical communication environments are complex and often continuously change, it is challenging to estimate these quantities. For example, although  $T_B$  can be calculated explicitly for non-adaptive algorithms (e.g., hierarchical search algorithm [6]), it depends on beam training strategies, external physical environments and desirable performance, which is difficult to obtain generally. To address this issue, we will propose an efficient algorithm based on ML in Section IV.

#### IV. BAYESIAN LEARNING TO ALLEVIATE BEAM DRIFT

To alleviate the BD effect, we propose an efficient algorithm based on Bayesian contextual bandit learning [27]–[29] in this section. The reader is referred to [30] for a complete exposition of contextual bandit learning and efficient learning algorithms (e.g., Bayesian posterior sampling or Thompson sampling).

##### A. Adaptive Design for Beam Training and Data Transmission

In this subsection, we propose a novel design philosophy for beam training (BT) and data transmission (DT) in mmwave communications. The design goal is to maximize the EAR performance  $\mathbb{E}(R_E)$ . Let  $C_k = \sqrt{-2k''(0)/k(0)}$ , which only depends on the communication environment. Let  $E$  represent the event that beam alignment succeeds. For an adaptive BT algorithm whose training overhead is not fixed, the expected EAR can be lower bounded by

$$\begin{aligned} \mathbb{E}(R_E) &\stackrel{(1)}{=} \mathbb{E}_{T_B E}(\mathbb{E}(R_E | T_B E)) \\ &\stackrel{(2)}{\geq} \mathbb{E}_{T_B E} \left( \left(1 - \frac{T_B}{T_S}\right) \left(1 - \frac{C_k}{2\pi} \exp\left(-\frac{b^2}{4k(0)}\right)\right) \cdot \log\left(1 + \frac{PN|g|^2 G_M}{\beta}\right) \right) \\ &\stackrel{(3)}{=} \left(1 - \frac{C_k}{2\pi} \exp\left(-\frac{b^2}{4k(0)}\right)\right) \cdot \log\left(1 + \frac{PN|g|^2 G_M}{\beta}\right) \mathbb{E}_{T_B E} \left(1 - \frac{T_B}{T_S}\right), \end{aligned} \quad (15)$$

where (1) is due to the conditional expectation formula, (2) is due to Theorem 2, and (3) is due to the fact that  $b$  and  $G_M$  are constants for a given codebook of data transmission. In general, the available training time resource  $T_B$  affects the probability of successful alignment and further the achievable rate. However, if the training resources (e.g., transmit power and available training time) are sufficient, beam alignment succeeds in a high probability. In this case,  $\mathbb{E}_{T_B E}(1 - T_B/T_S)$  can be approximated as  $1 - \mathbb{E}(T_B)/T_S$ . Then,  $\mathbb{E}(R_E)$  can be further simplified as

$$\mathbb{E}(R_E) \geq \left(1 - \frac{\mathbb{E}(T_B)}{T_S}\right) \left(1 - \frac{C_k}{2\pi} \exp\left(-\frac{b^2}{4k(0)}\right)\right) \cdot \log\left(1 + \frac{PN|g|^2 G_M}{\beta}\right). \quad (16)$$

It has been widely accepted that multi-resolution beams are more plausible for searching optimal beam directions (e.g., via hierarchical BT), while narrow beams are more suitable for DT, so as to achieve large array gains. However, the previous analysis indicates that the design methodology is problematic if the BD effect is taken into account. The reason is that narrow beams are vulnerable to the BD effect, which may deteriorate system performance. To maximize  $\mathbb{E}(R_E)$ , based on (16), we propose the following design philosophy:

- DT Design - mitigate the BD effect: In contrast to the conventional practice, we advocate that multi-resolution beams are employed for DT, and adapt the beam resolution or beam width depending on the rate of change of the environment. To achieve this goal, the design of DT is based on ML methodology.
- BT Design - minimize  $\mathbb{E}(T_B)$ : Narrow beams are used for BT, which is in contrast to the conventional practice. In general, the use of narrower beams often implies a larger training overhead. To minimize  $\mathbb{E}(T_B)$  and meanwhile guarantee successful beam alignment, the BT algorithm needs also to adapt the environment, i.e., it had better be a learning based algorithm.

**Remark 4.1** The rationale behind the BT design philosophy is locality principle of beam change (LPBC) - in practice the beam in the current time-slot is more likely to switch to one of its neighbor beams in the next time-slot, because AoA/AoD often changes continuously and the amplitude of the change is limited by maximal speed of the receiver/transmitter. Moreover, narrow beams often have high array gains, which help to combat the large path-loss in mmwave communications.

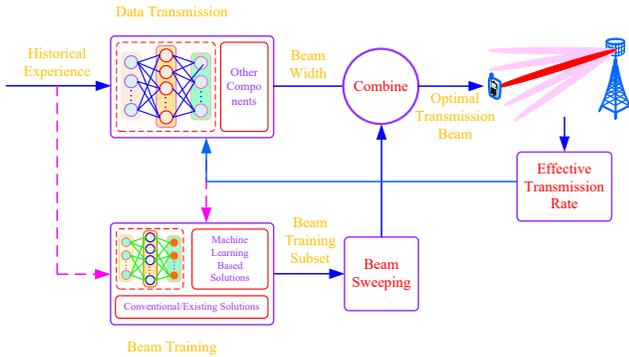


Fig. 4. The principle of the design of the BT and DT modules.

Based the proposed design philosophy of BT and DT, we introduce two modules referred to as BT module and DT module. In a broader sense, the BT and DT modules fulfill the functions CSI acquisition and precoder design. The sparsity of mmwave channels enables to acquire CSI via BT, and thus an arbitrary existing BT algorithm (e.g., hierarchical search or exhaustive search) can fulfil the function of the BT module. Besides alleviating the BD effect, the DT module also needs to fulfill other tasks, e.g., power allocation (or control) and interference elimination. As shown in Fig. 4, for the specific single-user scenario in this paper, the BT module is in charge of finding the optimal beam direction, with a minimum cost in terms of training overhead. The mission of the DT module

is to mitigate the BD effect (and allocate transmit power), so as to keep the SNR as large as possible and almost all the time in the phase of DT.

To achieve the best EAR performance, each of the two modules shall be designed based on the ML methodology, in order to sense the state of the environment (e.g., the rate of change of the environment) and further adapt the environment. To alleviate the burden of the UE, complex operations (e.g., to train the BT and DT modules, estimate beam direction and adjust beam width) are undertaken by the BS, while the UE only feeds back simple information (e.g., strengths of received signals) to the BS to help the BS to make decisions. Although the best option to realize the two modules is via the ML methodology, the DT module can also be embedded into the existing BT schemes which are not based on ML (e.g., the hierarchical search based algorithm) and helps to enhance the performance of these schemes. In the next subsection, we focus on the design of the DT module based on Bayesian bandit learning (BBL) [27]–[29]. The design of the BT module is similar, which is omitted due to space limitation. Note that designing and training the two modules separately to a large extent help to reduce computational complexity and exploit different characteristics of the two modules.

Before proceeding to details of the DT module design, we need to underscore the importance of distinguishing time-scales of optimizing, training or updating the two modules. We assume that in each time-slot, the BT module is trained or updated to generate a training beam subset used for sweeping beam space and further finding the optimal beam direction, i.e., the time-scale of updating the BT module is 1. However, if the time-scale of optimizing the DT module is still 1, the computational complexity of updating the two modules may be prohibitive, because each module may incorporate a neural network (NN) and requires intensive computation to train it.

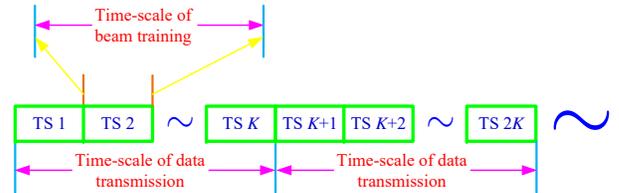


Fig. 5. Time-scales of optimizing the two modules - TS is short for time-slot. Note that the reciprocal of time-scale is update frequency. Hence, the frequency of training, updating and optimizing the BT module is  $K$ -times of that of the DT module.

To tackle this issue, as shown in Fig. 5, the time-scale of optimizing or updating the DT module is set to be larger than that of the BT counterpart, e.g., the DT module is optimized every  $K$  ( $K > 1$ ) time-slots. Moreover, by distinguishing the time-scales, another benefit is that more experiences/data can be accumulated when updating the DT module.

### B. Beam Width Optimization via Bayesian Bandit Learning

In this subsection, we utilize the BBL to design an efficient algorithm to optimize the beam width for DT. Note that in many cases optimal decisions depend on not only external

environments but also internal states of the system. However, there is no (or only one) state in conventional multi-armed bandit (MAB) model and decision making relies only on scalar rewards received from the environments, which inevitably leads to a large performance loss and even makes the algorithms fail. Hence, the MAB model is inapplicable here. In contrast, contextual bandit model incorporates both information of external environment and internal states of system into decision making, which enables the contextual bandit model to be applicable in complex systems and environments.

The core of employing BBL is to formulate the problem of beam width optimization (BWO) as a BBL problem, i.e., define prior and posterior distributions, action space, rewards and contexts. The role of BWO is to adjust the widths of beams used for data transmission via interacting with environments. The adjustment of beam width in this paper is implemented via multi-resolution codebooks. Therefore, there is a one-to-one correspondence between the actions and codebooks, and choosing an action is equivalent to choosing a codebook (for data transmission). First, based on collected empirical data, the BS chooses an action (i.e., beam width). Then, combining with beam direction information provided by the BT module, the beam used for data transmission can be determined. As more training data is collected, the chosen actions tend to be optimal.

Since the DT module is trained every  $K$  time-slots, to avoid confusion, the time range of every  $K$  time-slots is referred to as a time-unit and indexed by  $u$ , i.e., each time-unit consists of  $K$  time-slots. As mentioned earlier, to mitigate the BD effect, the DT module requires a multi-resolution codebook. Without loss of generality, a dyadic multi-resolution codebook  $\mathcal{C} = \{\mathcal{C}_1, \mathcal{C}_2, \dots, \mathcal{C}_S\}$ , which consists of  $S$  subcodebooks, is considered. Then, the action space is defined as

$$\mathcal{A} = \{1, 2, \dots, S\}.$$

Next, we define the rewards  $\{r_u\}$  and contexts  $\{X_u\}$ . The reward  $r_u$  corresponding to the  $u$ -th time-unit is defined by

$$r_u = \sum_{i=1}^K r_{u,i}, \quad (17)$$

where  $r_{u,i}$  represents the EAR corresponding to the  $i$ -th time-slot within the  $u$ -th time-unit. Let  $\bar{h}_{u,i}$ ,  $T_{u,i}^B$ ,  $P_{u,i}$  and  $G_{u,i}$  denote the corresponding equivalent channel coefficient, beam training overhead, transmit power and beamforming (or array) gain, respectively. The context  $X_u$ , which should reflect the system state, is defined as

$$X_u = (\bar{h}_{u,1}, \dots, \bar{h}_{u,K}, T_{u,1}^B, \dots, T_{u,K}^B, P_{u,1}, \dots, P_{u,K}, G_{u,1}, \dots, G_{u,K}). \quad (18)$$

To complete Bayesian bandit modeling, it is sufficient to assign posterior distribution models, which is, however, non-trivial. On the one hand, since communication environments are dynamic and complex, simple methods (e.g., linear algorithms) are insufficient to make wise decisions due to their lack of representational power [31]. On the other hand, although the NNs are sufficiently flexible, to provide estimations of

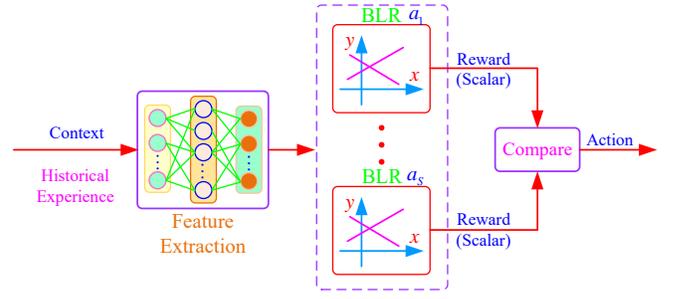


Fig. 6. The internal structure of the data transmission module. BLR is short for Bayesian linear regression.

uncertainty, the computational complexity may be prohibitive due to complex Bayesian posterior inference.

To address this dilemma, we incorporate NN and Bayesian linear regression, as shown in Fig. 6. Similar to [32], Bayesian linear regression is performed on top of the representation of the last layer of the NN [33]. Let  $r$  represent the reward. Then, the regression expression of  $r$  can be calculated as

$$r = \mathbf{w}^T f_X + \varepsilon, \quad (19)$$

where  $f_X = f(X)$  is the output of the NN for context  $X$  (as the input) and  $\varepsilon$  is distributed as  $\varepsilon \sim \mathcal{N}(0, \sigma^2)$ . To choose an appropriate action, we build a prediction model  $\mathbf{w}_a^T f(X) + \sigma_a^2$  for each action  $a \in \mathcal{A}$ , where  $\sigma_a^2$  is the power of noise  $\varepsilon_a$  of action  $a$ , i.e.,  $\varepsilon_a \sim \mathcal{N}(0, \sigma_a^2)$ . Note that the NN  $f(\cdot)$  is shared by all actions. Then, given a context  $X$ , we choose the action whose prediction reward is maximum, i.e.,  $a^*$  is chosen if  $\mathbf{w}_{a^*}^T f(X) + \sigma_{a^*}^2 \geq \mathbf{w}_a^T f(X) + \sigma_a^2 (\forall a \neq a^*)$ .

The use of Bayesian learning forces us to assign prior and posterior for the unknown parameters  $\{\mathbf{w}_a, \sigma_a^2 | a \in \mathcal{A}\}$ . To avoid intractability, the prior of  $(\mathbf{w}_a, \sigma_a^2)$  is modeled as

$$\pi_0^a(\mathbf{w}_a, \sigma_a^2) = \mathcal{N}(\mathbf{w}_a | \mathbf{m}_0, \sigma_a^2 \mathbf{S}_0) \cdot \text{IGam}(\sigma_a^2 | p_0, q_0), \quad (20)$$

where  $\text{IGam}(x | p, q)$  represents inverse Gamma distribution with shape parameter  $p$  and scale parameter  $q$ .<sup>4</sup> Note that  $\mathbf{m}_0$ ,  $\mathbf{S}_0$ ,  $p_0 > 0$  and  $q_0 > 0$  in (20) are constants and should be initialized appropriately. The reason for choosing the form in (20) (i.e., normal-inverse-gamma) is that each prior  $\pi_0^a$  is a conjugate prior, which facilitates posterior inference.

At time-unit  $u$ , after observing context  $X_u^a$  and reward  $r_u^a$  of the chosen action  $a$ , the posterior  $\pi_u^a$  is updated as

$$\pi_u^a(\mathbf{w}_a, \sigma_a^2) = \mathcal{N}(\mathbf{w}_a | \mathbf{m}_u^a, \sigma_a^2 \mathbf{S}_u^a) \cdot \text{IGam}(\sigma_a^2 | p_u^a, q_u^a). \quad (21)$$

The parameters  $\mathbf{m}_u^a$ ,  $\mathbf{S}_u^a$ ,  $p_u^a$  and  $q_u^a$  are given by [31]

$$\begin{aligned} \mathbf{S}_u^a &= (\mathbf{S}_0^{-1} + \mathbf{X}_u^a \mathbf{X}_u^a)^{-1} \\ \mathbf{m}_u^a &= \mathbf{S}_u^a (\mathbf{S}_0^{-1} \mathbf{m}_0 + \mathbf{X}_u^a \mathbf{r}_u^a) \\ p_u^a &= p_0 + L_u/2 \\ q_u^a &= q_0 + \frac{\mathbf{m}_0^T \mathbf{S}_0^{-1} \mathbf{m}_0 - (\mathbf{m}_u^a)^T (\mathbf{S}_u^a)^{-1} \mathbf{m}_u^a + \mathbf{r}_u^a \mathbf{r}_u^a}{2}, \end{aligned} \quad (22)$$

<sup>4</sup>The probability density function of  $\text{IGam}(x | p, q)$  is given by

$$h(x | p, q) = \frac{q^p}{\Gamma(p)} x^{-(p+1)} e^{-q/x}.$$

where  $\mathbf{X}_a = [f_{X_{(1)}^a}, \dots, f_{X_{(L_a)}^a}]$  and  $\mathbf{r}_a = [r_{(1)}^a, \dots, r_{(L_a)}^a]$  collect accumulated experiences of action  $a$  with  $L_a$  denoting the number of samples of action  $a$  until time-unit  $u$  and  $f_{X_{(i)}^a}$  and  $r_{(i)}^a$  denoting the  $i$ -th sample.

Next, we proceed to the training of the NN. The training dataset, e.g., for action  $a \in \mathcal{A}$ , obtained via interacting with the environment continuously, is given by

$$\{(X_{(1)}^a, r_{(1)}^a), \dots, (X_{(L_a)}^a, r_{(L_a)}^a) \mid a \in \mathcal{A}\}. \quad (23)$$

Note, however, that this training dataset cannot be used to train the NN directly, because a training dataset which can be used to train the NN directly should take the form  $\{(X_{(i)}^a, y_{(i)}^a = f(X_{(i)}^a)) \mid i = 1, \dots, L_a\}$ . To tackle this issue, in view that  $r$  is linear with respect to  $f_X$  in (19), a linear layer with identity activation function (i.e.,  $\sigma(x) = x$ ) is firstly appended to the NN. Then, the training dataset in (23) can be used to train the augmented network. After finishing the training of the augmented network, the original NN has already been trained as well. When the (original) NN is trained, the samples used to update the posterior in (21) have to be updated based on the newly trained NN. More specifically, the samples  $\mathbf{X}_a = [f_{X_{(1)}^a}, \dots, f_{X_{(L_a)}^a}]$  are replaced by  $\mathbf{X}_a = [f'_{X_{(1)}^a}, \dots, f'_{X_{(L_a)}^a}]$ , with  $f'$  denoting the newly trained NN.

**Remark 4.2** To further reduce the computational complexity of updating the NN, the linear regression model and the NN can be updated at different time-scales, e.g., the time-scale of updating the NN is  $KF$  with  $F(> 1)$  a positive integer.

---



---

**Algorithm 1:** Intelligent Beam Width Optimization via BBL

---

1: **input:** multi-resolution codebook  $\mathcal{C} = \{\mathcal{C}_1, \mathcal{C}_2, \dots, \mathcal{C}_S\}$ ; action space  $\mathcal{A} = \{1, 2, \dots, S\}$ ; update frequency or time-scale of NN  $F$

---

2: **initialize** prior  $\{\pi_0^a \mid a \in \mathcal{A}\}$  (i.e.,  $\mathbf{m}_0, \mathbf{S}_0, p_0$  and  $q_0$ ); parameters of NN; memories  $\{\mathcal{D}_a = \emptyset\}$ ; DT sub-codebook  $\mathcal{C}_1$ ; let  $u = 1$

---

3: **repeat** for each time-unit

- (a) construct context  $X_u$  according to (18)
- (b) compute nonlinear feature  $f_{X_u}$  for  $X_u$  via NN
- (c) sample  $\pi_u^a$  for each action  $a \implies \{\mathbf{w}_{a,u}, \sigma_{a,u}^2\}$
- (d) determine optimal action  $a_u^*$  with  $a_u^*$  given by
 
$$a_u^* = \arg \max_{a \in \mathcal{A}} \{\mathbf{w}_{a,u}^\top f_{X_u} + \sigma_{a,u}^2\}$$
- (e) **perform**  $a_u^* \implies$  **switch to DT subcodebook**  $\mathcal{C}_{a_u^*}$
- (f) receive reward  $r_u$  (from external environment)
- (g) update  $\pi_u^a$  to  $\pi_{u+1}^a$  according to (22)
- (h) update memory:  $\mathcal{D}_{a_u^*} \leftarrow \mathcal{D}_{a_u^*} \cup \{(X_u, a_u^*, r_u)\}$
- (i) if  $u \bmod F = 0$ , update NN via back propagation
- (j) let  $u \leftarrow u + 1$

**end**

---



---

For clarity, the designed BWO algorithm is summarized in Algorithm 1. For a given multi-resolution codebook  $\mathcal{C} = \{\mathcal{C}_1, \mathcal{C}_2, \dots, \mathcal{C}_S\}$ , we construct action space  $\mathcal{A} = \{1, \dots, S\}$ . Then, we initialize the priors of all actions and parameters (i.e., weights and biases) of the underlying NN. The experience memory of each action is emptied and the time-unit counter is set to 1. The codebook used for DT is initialized with an arbitrary subcodebook, e.g.,  $\mathcal{C}_1$ . Next, we repeat the operations from (a) to (j) in each time-unit. First, the context  $X_u$  is

constructed based on the information collected within time-unit  $u$ . In step (b), with  $X_u$  available, the feature  $f_{X_u}$  is extracted via forward propagation (FP) of the NN. In step (c), parameters  $\mathbf{w}_{a,u}$  and  $\sigma_{a,u}^2$  of each action  $a$  are obtained by sampling the posterior  $\pi_u^a$ . With the optimal action  $a_u^*$  determined in step (d), we perform the action by switching the DT (sub-)codebook to  $\mathcal{C}_{a_u^*}$  in step (e). At the end of each time-unit, reward  $r_u$  is received in step (f). The posterior and memory are updated in steps (g) and (h), respectively. In step (i), since sufficient experiences have been accumulated, the NN should be updated via back propagation (BP). At the end of each time-unit, the counter is increased by 1.

Next, we consider issues related to algorithm implementation. The first one is storage resource requirement (i.e., to cache  $\mathcal{D}$ ). Since each sample in  $\mathcal{D}$  takes a simple form  $\{(X, a, r)\}$  (with  $X, a$  and  $r$  data of small size) and outdated samples can be safely removed from  $\mathcal{D}$  (thanks to the online implementation manner), very few storage resources are required. Another one is computational issue, which is mainly caused by NN operations, i.e., FP and BP. In general, BP requires intensive computations, which, however, can be accomplished in the phase of data transmission, and thus it does not affect real-time implementation. The computational complexity of FP can almost be negligible. The typical fully-connected NN is taken as an example. With the assumption that the NN consists of  $L + 1$  layers with the  $l$ -th layer having  $n_l$  neurons and the ReLU is chosen as the activation function, the computational complexity is dominated by  $\sum_{l=1}^L n_{l-1}n_l$  multiplication operations and  $\sum_{l=1}^L n_l$  addition operations. Therefore, compared to other algorithms, Algorithm 1 has an appealing and competitive advantage on real-time implementation. In particular, it does not involve complex calculations (e.g., matrix inversion) or time-consuming iterations.

Finally, we clarify how to coordinate or integrate the BT and DT modules. The BT module provides beam direction in each time-slot as accurate as possible, and the DT module chooses a DT (sub-)codebook that can achieve the best EAR performance in each time-unit. The DT codebook chosen by the DT module is denoted by  $\mathcal{C}_a = \{\mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_{N_a}\}$  of size  $N_a$ . The beam center angle of each beam  $\mathbf{f}_i \in \mathcal{C}_a$  is denoted by  $\phi_i$ , which often corresponds to beam direction of highest array gain. Let  $\phi_0$  represent the beam direction found by the BT module (or an arbitrary existing BT algorithm). Then, the beam used for DT is given by

$$\mathbf{f}_{i^*} \in \mathcal{C}_a \quad \text{with} \quad i^* = \arg \min_{i \in [N_a]} |\phi_0 - \phi_i|. \quad (24)$$

## V. SIMULATION RESULTS

In this section, simulation results are provided to demonstrate the performance of the proposed algorithms. The number of antennas of ULA is  $N = 64$ . Without loss of generality, a dyadic multi-resolution codebook is considered. As shown in Fig. 7, the codebook  $\mathcal{C}$  consists of 6 subcodebooks, i.e.,  $\mathcal{C}_1, \mathcal{C}_2, \dots, \mathcal{C}_6$ , whose sizes are 2, 4, 8, 16, 32 and 64, respectively. For convenience, each subcodebook  $\mathcal{C}_i$  ( $i = 1, \dots, 6$ ) is abbreviated as SCB $i$ . The beam width corresponding to  $\mathcal{C}_i$  is  $2^{1-i}$ , i.e., the beam width decreases as  $i$  increases. In

the following experiments, only SCB3 - SCB6 are utilized to evaluate different algorithms. The reason is that the array gains of the beams in SCB1 and SCB2 are too small to achieve a good EAR performance. The metrics probability of successful alignment (PSA) and EAR are adopted to evaluate different algorithms. PSA is defined as the ratio between the number of successful alignments and that of total alignments, which can be estimated via Monte-Carlo simulations.

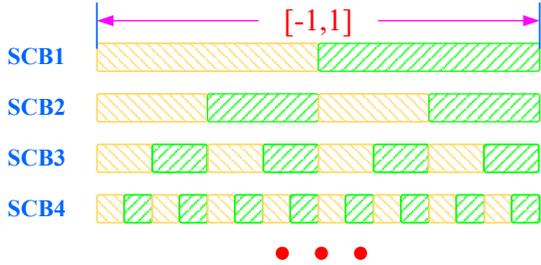


Fig. 7. The structure of the dyadic hierarchical codebook.

For all simulation experiments, the channel model in (1) includes one LOS path and three NLOS paths. The AoDs of the three NLOS paths are uniformly distributed in  $[0, 2\pi]$ . The average power ratio between the LOS path gain  $g_{\text{LOS}}$  and each NLOS path gain  $g_{\text{NLOS}}$  is 10dB. The path gain of each NLOS path  $g_{\text{NLOS}}$  is distributed as  $\mathcal{CN}(0, \sigma_{\text{NLOS}}^2)$ , where  $\sigma_{\text{NLOS}}^2$  is calculated according to the path gain of the LOS path. To accommodate practical complex environments, in particular, sufficient randomness, GPs with the SM kernel can be used to simulate the change of directions of (physical) channel paths. The parameters  $\{\mu_q | q = 1, \dots, Q\}$  in (7) characterize AoD or AoA change of the LOS path. In this case, let  $Q = 1$  and denote  $\mu_1$  as  $\mu$ . More simply, the sine function with  $\mu$  denoting the frequency can also be used to simulate AoD/AoA change of the LOS path. A larger value of  $\mu$  implies that AoD or AoA changes more rapidly. The fully-connected NN with 2 layers (and 16 and 8 neurons in the two layers) and the ReLU activation function are considered in Algorithm 1. Extensive simulation experiments show that Algorithm 1 also works well for other network settings, e.g., the convolutional/recurrent NN, the Sigmoid function, the number of layers 2 ~ 6, and the number of neurons per layer 8 ~ 64.

First, we confirm the effectiveness and superiority of the BT design philosophy, in particular, LPBC. In the following results we compare our approach against the closest relevant benchmarks. Specifically, the following methods are evaluated:

- EXH - exhaustive search based BT algorithm [5]. All beams in the codebook are used to search the optimal beam direction. Hence, the training overhead of EXH is equal to the size of the codebook, which is the highest among the four BT algorithms.
- HIE - conventional hierarchical search based BT algorithm [7]. Low-resolution or wide beams are tried first, and then high-resolution or narrow beams are employed. The training overhead of HIE is fixed to  $2 \log |\mathcal{C}|$  with  $|\mathcal{C}|$  denoting the size of the codebook.
- MAB - multi-armed bandit based BT algorithm [19].

Each beam in the codebook is regarded as an arm, and a well-known upper confidence bound (UCB) algorithm [28] is chosen to select arms.

- SBL - stochastic bandit learning based BT algorithm [34]. SBL is a standard BT algorithm designed based on LPBC. To implement LPBC, the arms or actions are constructed based on beam index difference technique [34]. SCB5 and SCB6 are considered only in the experiments.

Fig. 8 shows the PSA performance of different BT algorithms.

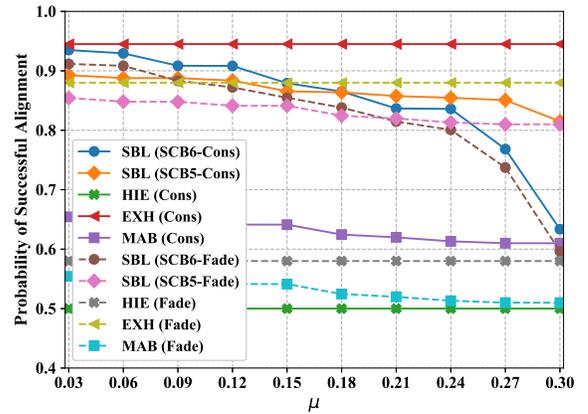


Fig. 8. The PSA performance of different BT algorithms (with different subcodebooks) - SNR = 5dB. For “SCB $i$ -Cons” and “SCB $i$ -Fade” ( $i = 5$  or 6), the same subcodebook SCB $i$  is chosen, while the distributions of the LOS path gain  $g_{\text{LOS}}$  are different. Specifically, “CONS” and “FADE” represent  $|g_{\text{LOS}}| = 1$  and  $g_{\text{LOS}} \sim \mathcal{CN}(m, 1)$  with  $|m| = 1$ , respectively.

It is seen from Fig. 8 that the SBL algorithm achieves a better performance than the HIE algorithm and approaches that of the EXH algorithm. Moreover, for the case of fading LOS paths, i.e.,  $g_{\text{LOS}} \sim \mathcal{CN}(m, 1)$  with  $|m| = 1$ , the SBL algorithm even performs better than the EXH algorithm.<sup>5</sup> Interestingly, the SBL algorithm with SCB5 that combines LPBC with simple beam search<sup>6</sup> also achieves a good performance - a bit worse than SBL with SCB6 when  $\mu$  is less than (about) 0.2 and better than it when  $\mu$  becomes larger. The reason for this is that for the same size of training beam subset, the coverage of the subset chosen from SCB5 is larger than that of the subset chosen from SCB6. In fact, SBL with SCB5 enjoys the advantages of low training overhead (due to LPBC) and high array gain (due to narrow beams). Without special

<sup>5</sup>The reason for this is as follows. The set of beams used for training is denoted by  $\mathcal{T} \subset \mathcal{C}$ . If beam  $\mathbf{f}_i \in \mathcal{T}$  is chosen, the received signal  $y_i \in \mathbb{C}$  is given by  $y_i = \sqrt{P} \mathbf{h}^H \mathbf{f}_i s + n_i$ . Without loss of generality, let  $\mathbf{f}_{i^*}$  be the optimal beam. Note that the beam  $\mathbf{f}_{i^*}$  can be identified if and only if

$$Y_{i^*} = |y_{i^*}|^2 > Y_i = |y_i|^2, (\forall i \neq i^*, i \in \mathcal{T}). \quad (25)$$

Intuitively, if more variables  $\{Y_i\}$  are involved in the comparison in (25) (accordingly, more beams are used for training), the probability that the events in (25) occur simultaneously becomes smaller. Because the beam training subset used by the EXH algorithm is much smaller than that used by the SBL algorithm. Therefore, the impact of noise on the SBL algorithm is much smaller than that on the EXH algorithm. In the case of fading channel path gain,  $|\mathbf{h}^H \mathbf{f}_i|$  may be very small and thus the impact of noise becomes more salient. As a result, the SBL algorithm performs better than the EXH algorithm in this case.

<sup>6</sup>Note that the beam width of SCB5 is larger than that of SCB6, to achieve the same accuracy of beam alignment, a simple beam search is required.

treatments (e.g., the beam index difference technique [34]), MAB mainly applies to slow-varying channel environments. This is the reason why SBL performs better than MAB. All the observations demonstrate the effectiveness and superiority of the BT design philosophy.

Next, we demonstrate the average EAR performance under different simulation settings. The BT module is implemented via the SBL algorithm that provides the optimal beam direction, and the DT module is implemented via Algorithm 1 that optimizes the beam width. Fig. 9 shows the EAR performance with 4 subcodebooks, i.e., SCB3, SCB4, SCB5 and SCB6. It is observed that as the environment changes faster and faster and thus the impact of the BD effect becomes more serious, the EAR performances achieved with different subcodebooks decrease accordingly. However, the EAR performance of BWO is larger than that of any fixed subcodebook. Moreover, the performance gap between them becomes larger and larger, as the environment changes faster and faster. These observations indicate that the designed BWO algorithm can effectively mitigate the BD effect and guarantee a good EAR performance in the dynamic channel environment.

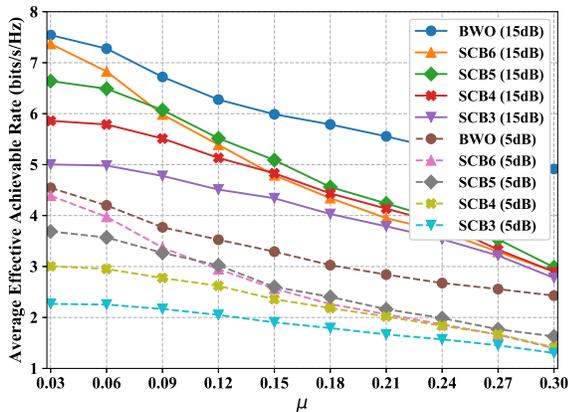


Fig. 9. The average EAR performance vs the rate of change of the environment with different subcodebooks (or equivalently, beam widths) -  $|g_{LOS}| = 1$ .

Interestingly, it can be observed that when the environment changes relatively slow (e.g., when  $\mu < 0.1$  for the setting of 5dB), the EAR performance achieved with SCB6 is the highest among the 4 subcodebooks. The reason for this is that in this case the BD effect is not significant, and meanwhile the array gains of SCB6 are the highest due to the most narrow beam width. As the BD effect becomes more and more significant, the EAR performance achieved with SCB6 deteriorates and becomes worse than that with SCB5, which shows that narrow beams are vulnerable to the BD effect. It can also be observed that the EAR performance achieved with SCB3 is the worst in the two SNR settings. The reason for this is that although wide beams are robust to the BD effect, their array gains are relatively low, which, in turn, reduces the EAR performance. In contrast, our algorithm can achieve a good balance between achieving high array gains and mitigating the BD effect, and sense and adapt the changing environment in real-time.

The EAR performance vs SNR is shown in Fig. 10. One

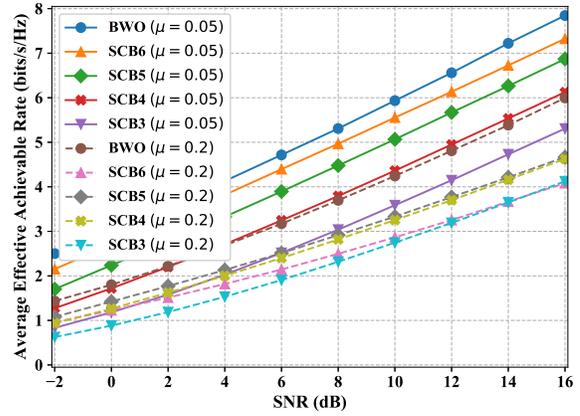


Fig. 10. The average EAR performance vs SNR (with different subcodebooks) -  $|g_{LOS}| = 1$ .

can observe that as the SNR increases, the EAR performance achieved with different subcodebooks also increases. However, for the two environment settings (i.e.,  $\mu = 0.05$  and  $\mu = 0.20$ ), the EAR performance achieved with BWO is much better than that with any fixed subcodebook. The reason for this is that the designed BWO algorithm can sense the rate of change of the environment and adjust the beam width in real-time, which achieves a good tradeoff between obtaining high array gains and mitigating the BD effect. One can also observe that for a fast changing environment, a subcodebook of narrow beams (e.g., SCB6) is more preferable in the low SNR region, while a subcodebook of wide beams achieves a better performance in the high SNR region. The reason for this is that in the low SNR region, a high array gain plays a more important role in improving the performance in terms of achievable rate.

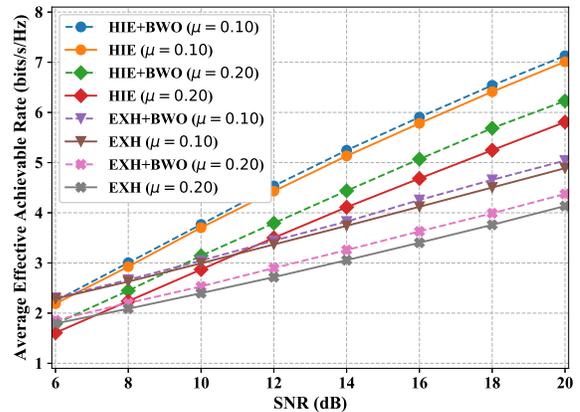


Fig. 11. The average EAR performance vs SNR (with and without beam width optimization) -  $|g_{LOS}| = 1$ .

An important advantage of the proposed BWO algorithm is that it can be flexibly embedded into a variety of the existing DT algorithms (i.e., without changing the architecture or communication protocol of the existing wireless communication systems) and helps to improve their EAR performance. The HIE and EXH algorithms are taken as two examples. Note that once the optimal beam (direction) is found by the HIE or EXH

algorithm, it can also be used to transmit data, i.e., the HIE or EXH algorithm can also be regarded as a DT algorithm. Specifically, for the HIE algorithm, the codebook SCB6 is used for both BT and DT (i.e., multi-resolution beams for BT and narrow beams for DT), while for the EXH algorithm, BT and DT share the same codebook. The EAR performance of the two algorithms with and without BWO is shown in Fig. 11. It can be observed that the BWO algorithm can improve the EAR performance of the two algorithms, and the improvement is more significant if the environment changes faster.

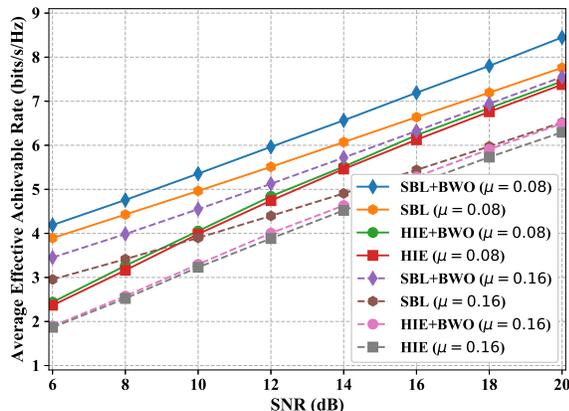


Fig. 12. The average EAR performance of different algorithms with various types of beams for BT and DT.

TABLE I  
ALGORITHMS AND TYPES OF BEAMS FOR BT AND DT

Algorithm	Type of Beam for BT	Type of Beam for DT
HIE	Multi-Resolution Beams	Narrow Beams
HIE+BWO	Multi-Resolution Beams	Multi-Resolution Beams
SBL	Narrow Beams + LPBC	Narrow Beams
SBL+BWO	Narrow Beams + LPBC	Multi-Resolution Beams

Finally, to further verify the effectiveness and superiority of the proposed design philosophy, EAR performance of different algorithms with various types of beams for BT and DT is shown in Fig. 12. The details of these algorithms are provided in Table I, where “SBL” (with or without BWO) indicates that the algorithms use narrow beams for BT along with LPBC. It is seen that an algorithm using multi-resolution beams for DT and narrow beams (along with LPBC) for BT achieves the best performance, which demonstrates the effectiveness and superiority of the design philosophy. The reason for this is that narrow beams which have large array gains help to improve BT efficiency, and meanwhile multi-resolution beams guarantee a large EAR performance by alleviating the BD effect.

## VI. CONCLUSION

In this paper, we revealed and studied the BD phenomenon caused by non-ideal features inherent in practical beams and rapidly changing environments, which has a negative impact on system performance. To mitigate the BD effect, we first theoretically analyzed the OP and EAR performance, which took practical factors (e.g., beam width, transmit power and

the rate of change of environments) into consideration. Then, we proposed a novel DT and BT design philosophy, where a multi-resolution codebook is employed for DT while narrow beams are suggested for BT. Finally, to mitigate the BD effect, we proposed a learning based BWO algorithm. Comprehensive simulation results were provided to demonstrate the effectiveness and superiority of the designed algorithms.

## APPENDIX A PROOF OF THEOREM 1

Since the considered GP is stationary, it is sufficient to focus on the typical interval  $[0, I]$  via translation, where  $I = T_S - T_B$  is the time of effective data transmission. To proceed, we introduce the concept of level crossing. The motivation for this is that the main-lobe of a beam is often an interval and the occurrence of outage is closely related to crossing endpoints of the interval. For a sample function  $\{x(t)\}$  of a continuous process  $\{X(t), t \in \mathbb{R}\}$ ,  $x(t)$  has an up-crossing of level  $u$  at  $t_0$  if, for some  $\varepsilon > 0$ ,  $x(t) \leq u$  for all  $t \in (t_0 - \varepsilon, t_0]$  and  $x(t) \geq u$  for all  $t \in [t_0, t_0 + \varepsilon)$ . For an interval  $\mathcal{I}$ ,  $N_u^+(\mathcal{I})$  represents the number of up-crossings by  $x(t)$  in  $\mathcal{I}$ . Similarly,  $N_u(\mathcal{I})$  represents the number of crossings, i.e., the number of  $t \in \mathcal{I}$  such that  $x(t) = u$ .

The intensity of the up-crossings is a non-negative function  $\mu_u^+(t)$  such that

$$\int_{t \in \mathcal{I}'} \mu_u^+(t) dt = \mathbb{E}(N_u^+(\mathcal{I}')), \quad \text{for every interval } \mathcal{I}'.$$

The famous Rice’s formula characterizes the intensity of up-crossings [35]. The Rice formula of a GP version is provided in the following lemma.

**Lemma 1.** *Let  $N_u$  be the number of up-crossings of the level  $u$  by a zero-mean, stationary, almost surely continuous Gaussian process on  $[0, 1]$ , then*

$$\mathbb{E}(N_u) = \frac{\lambda_2^{1/2}}{2\pi\sigma} \exp\left(-\frac{u^2}{2\sigma^2}\right), \quad (26)$$

where  $\sigma^2 = \mathbb{E}(|X(t)|^2) = k(0)$  and  $\lambda_2 = -k''(0)$ .

Before proceeding to the formal proof of Theorem 1, we shall first explain  $k(0)$  and  $k''(0)$  of  $X(t)$ . In fact,  $k(0)$  characterizes the variance of  $X(t)$  at an arbitrary time. Next, we explain  $k''(0)$  from the perspectives of both time-domain and frequency-domain. The Bochner’s theorem (Theorem 3.3 in [26]) asserts that a stationary stochastic process is associated with a spectral distribution function, denoted by  $F(\omega)$ , such that the spectral distribution function and the correlation function compose a Fourier transform pair.

**Time-domain Interpretation:** Let  $X'(t)$  be the derivative process of  $X(t)$ . Theorem 2.2 in [26] shows  $-k''(0) = \mathbb{E}[X'(t)X'(t)]$ , i.e.,  $-k''(0)$  is the variance of the derivative process  $X'(t)$ .

**Frequency-domain Interpretation:** Theorem 2.2 in [26] also shows  $-k''(0) = \int_{-\infty}^{\infty} \omega^2 dF(\omega)$ , i.e.,  $-k''(0)$  is equal to the second-order spectral moment. If  $-k''(0)$  is large, more high-frequency components are involved, i.e., the corresponding GP

changes rapidly.<sup>7</sup>

Now, we proceed to prove Theorem 1. Based on Lemma 1, we can bound  $P_{\text{out}}(T) = \mathbb{P}(\text{SNR} < T)$ . In practice, it may be challenging to directly calculate a desired probability or expectation. A useful trick is to take condition (i.e., to first “condition” on some appropriate random variable or element). Specifically, let  $p_0(x)$  represent the probability density of  $f_0 = f(0)$ , i.e.,

$$p_0(x) = \frac{1}{\sqrt{2\pi k(0)}} \exp\left(-\frac{x^2}{2k(0)}\right).$$

Then,  $P_{\text{out}}(T)$  can be rewritten as

$$P_{\text{out}}(T) = \int_{\mathbb{R}} p_0(x) \mathbb{P}(\text{SNR} < T | f_0 = x) dx. \quad (27)$$

Note that if  $PN|g|^2\beta^{-1}G_M < T$ ,  $\mathbb{P}(\text{SNR} < T | f_0 = x) = 1$ , and thus  $P_{\text{out}}(T) = 1$ . Hence, it is sufficient to consider the case  $PN|g|^2\beta^{-1}G_M \geq T$ .

Let  $\mathcal{I} = [0, T]$  for simplicity. The conditional probability  $\mathbb{P}(\text{SNR} < T | f_0 = x)$  can be calculated as

$$\begin{aligned} & \mathbb{P}(\text{SNR} < T | f_0 = x) \\ &= \mathbb{P}(f(t) \geq f_0 + b \cup f(t) \leq f_0 - b | f_0 = x, t \in \mathcal{I}) \\ &= \mathbb{P}(f(t) \geq f_0 + b | f_0 = x, t \in \mathcal{I}) + \\ & \quad \mathbb{P}(f(t) \leq f_0 - b | f_0 = x, t \in \mathcal{I}). \end{aligned}$$

$\mathbb{P}(f(t) \geq f_0 + b | f_0 = x, t \in \mathcal{I})$  can be calculated as

$$\begin{aligned} & \mathbb{P}(f(t) \geq f_0 + b | f_0 = x, t \in \mathcal{I}) \\ &= \mathbb{P}(N_{f_0+b}(\mathcal{I}) \geq 1) \stackrel{(1)}{\leq} \mathbb{E}(N_{f_0+b}(\mathcal{I})) \stackrel{(2)}{=} I \mathbb{E}(N_{f_0+b}) \\ & \stackrel{(3)}{=} \frac{I}{2\pi} \sqrt{\frac{-k''(0)}{k(0)}} \exp\left(-\frac{(f_0 + b)^2}{2k(0)}\right), \end{aligned}$$

where (1) is due to the Markov’s inequality, (2) is due to the stationary property of the GP, and (3) is obtained by using the Rice’s formula in Lemma 1. The use of the Markov’s inequality is important, since, as a bridge, it links the probability and expectation and, more importantly, facilitates the use of Lemma 1. Note that if  $T_S$  is not large, (1) is, in fact, an equality. In fact, in this case the number of possible up-crossings is almost always at most 1.

$\mathbb{P}(f(t) \leq f_0 - b | f_0 = x, t \in \mathcal{I})$  can be calculated similarly, which is given by

$$\begin{aligned} & \mathbb{P}(f(t) \leq f_0 - b | f_0 = x, t \in \mathcal{I}) \\ & \leq \frac{I}{2\pi} \sqrt{\frac{-k''(0)}{k(0)}} \exp\left(-\frac{(f_0 - b)^2}{2k(0)}\right). \end{aligned}$$

Then,  $\mathbb{P}(\text{SNR} < T | f_0 = x, t \in \mathcal{I})$  can be rewritten as

$$\begin{aligned} & \mathbb{P}(\text{SNR} < T | f_0 = x, t \in \mathcal{I}) \\ & \leq \frac{I}{2\pi} \sqrt{\frac{-k''(0)}{k(0)}} \exp\left(-\frac{b^2 + f_0^2}{2k(0)}\right) \\ & \quad \left(\exp(-bf_0/k(0)) + \exp(bf_0/k(0))\right). \quad (28) \end{aligned}$$

<sup>7</sup>An example is shown in Fig. 3, where the kernel is the SE kernel  $k_{\text{SE}}(t, t') = \sigma_f^2 \exp\left(-\frac{1}{2l^2}|t - t'|^2\right)$  with  $-k''(0) = \sigma_f^2 l^{-2}$ . A small value of parameter  $l$  implies that the GP changes rapidly, as shown in Fig. 3.

By plugging (28) into (27) and via some algebraic operations, we can complete the proof of this theorem. The subsequent derivation is straightforward, which is omitted.

## APPENDIX B PROOF OF THEOREM 2

Note that for a non-negative continuous random variable  $X$ , its mathematical expectation can be calculated as

$$\mathbb{E}(X) = \int_0^\infty \mathbb{P}(X \geq t) dt. \quad (29)$$

Let  $(\Omega, \mathcal{F}, \mathbb{P})$  be the underlying probability space. Then, the expectation of  $X$  can be calculated as

$$\begin{aligned} \mathbb{E}(X) &= \int_{\Omega} X(\omega) d\mathbb{P}(\omega) = \int_{\Omega} \int_0^\infty \mathbb{I}_{X(\omega) \geq t} d\mathbb{P}(\omega) dt \\ & \stackrel{(*)}{=} \int_0^\infty dt \int_{\Omega} \mathbb{I}_{X(\omega) \geq t} d\mathbb{P}(\omega) = \int_0^\infty \mathbb{P}(X \geq t) dt, \end{aligned}$$

where (\*) is due to the Fubini-Tonelli theorem.

Under the assumption that the overhead of beam training within each time-slot is fixed, according to (29), the expected EAR performance can be calculated as

$$\begin{aligned} \mathbb{E}(R_E) &= T_S^{-1} \mathbb{E}\left(\int_{T_B}^{T_S} \log(1 + \text{SNR}(t)) dt\right) \\ &= T_S^{-1} \int_{T_B}^{T_S} \mathbb{E}\left(\log(1 + \text{SNR}(t))\right) dt \\ &= T_S^{-1} \int_0^\infty dx \int_{T_B}^{T_S} \mathbb{P}(\text{SNR}(t) \geq e^x - 1) dt \quad (30) \end{aligned}$$

Note that  $\mathbb{P}(\text{SNR}(t) \geq e^x - 1) = 0$  if  $x > \log(1 + PN|g|^2 \cdot \beta^{-1}G_M)$ , and otherwise,  $\mathbb{P}(\text{SNR}(t) \geq e^x - 1) = 1 - \mathbb{P}(\text{SNR}(t) < e^x - 1)$  with  $\mathbb{P}(\text{SNR}(t) < e^x - 1)$  calculated in (12). Let  $x_0 = \log(1 + PN|g|^2\beta^{-1}G_M)$ . Then,  $\mathbb{E}(R_E)$  can be further calculated as

$$\begin{aligned} \mathbb{E}(R_E) &= T_S^{-1} \int_0^{x_0} dx \int_{T_B}^{T_S} \left(1 - \mathbb{P}(\text{SNR}(t) < e^x - 1)\right) dt \\ &= \left(1 - \frac{T_B}{T_S}\right) \left(1 - \frac{1}{2\pi} \sqrt{\frac{-2k''(0)}{k(0)}} \exp\left(-\frac{b^2}{4k(0)}\right)\right) \\ & \quad \log\left(1 + PN|g|^2\beta^{-1}G_M\right), \quad (31) \end{aligned}$$

which completes the proof.

## REFERENCES

- [1] M. Xiao, S. Mumtaz, Y. Huang, L. Dai, Y. Li, M. Matthaiou, G. K. Karagiannidis, E. Björnson, K. Yang, C. L. I, and A. Ghosh, “Millimeter wave communications for future mobile networks,” *IEEE J. Sel. Areas Commun.*, vol. 35, no. 9, pp. 1909–1935, Sept 2017.
- [2] X. Gao, L. Dai, Z. Chen, Z. Wang, and Z. Zhang, “Near-optimal beam selection for beamspace mmwave massive MIMO systems,” *IEEE Commun. Lett.*, vol. 20, no. 5, pp. 1054–1057, May 2016.
- [3] J. Singh and S. Ramakrishna, “On the feasibility of codebook-based beamforming in millimeter wave systems with multiple antenna arrays,” *IEEE Trans. Wireless Commun.*, vol. 14, no. 5, pp. 2670–2683, May 2015.
- [4] Z. Xiao, T. He, P. Xia, and X. G. Xia, “Hierarchical codebook design for beamforming training in millimeter-wave communication,” *IEEE Trans. Wireless Commun.*, vol. 15, no. 5, pp. 3380–3392, May 2016.

- [5] S. He, J. Wang, Y. Huang, B. Ottersten, and W. Hong, "Codebook-based hybrid precoding for millimeter wave multiuser systems," *IEEE Trans. Signal Process.*, vol. 65, no. 20, pp. 5289–5304, Oct 2017.
- [6] 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.
- [7] S. Hur, T. Kim, D. Love, J. Krogmeier, T. Thomas, and A. Ghosh, "Millimeter wave beamforming for wireless backhaul and access in small cell networks," *IEEE Trans. Commun.*, vol. 61, no. 10, pp. 4391–4403, October 2013.
- [8] J. Song, J. Choi, and D. J. Love, "Common codebook millimeter wave beam design: Designing beams for both sounding and communication with uniform planar arrays," *IEEE Trans. Commun.*, vol. 65, no. 4, pp. 1859–1872, April 2017.
- [9] D. Zhu, J. Choi, C. Cheng, W. Xiao, and R. W. Heath, "High-resolution angle tracking for mobile wideband millimeter-wave systems with antenna array calibration," *IEEE Trans. Wireless Commun.*, vol. 17, no. 11, pp. 7173–7189, 2018.
- [10] D. Zhang, H. Chen, M. Shirvanimoghaddam, Y. Li, and B. Vucetic, "Training beam sequence optimization for millimeter wave mimo tracking systems," in *2018 IEEE ICC*, May 2018, pp. 1–6.
- [11] V. Va, H. Vikalo, and R. W. Heath, "Beam tracking for mobile millimeter wave communication systems," in *2016 IEEE GlobSIP*, Dec 2016, pp. 743–747.
- [12] A. Alkhatieb, S. Alex, P. Varkey, Y. Li, Q. Qu, and D. Tujkovic, "Deep learning coordinated beamforming for highly-mobile millimeter wave systems," *IEEE Access*, vol. 6, pp. 37328–37348, 2018.
- [13] V. Va, J. Choi, T. Shimizu, G. Bansal, and R. W. Heath, "Inverse multipath fingerprinting for millimeter wave V2I beam alignment," *IEEE Trans. Veh. Technol.*, vol. 67, no. 5, pp. 4042–4058, May 2018.
- [14] J. C. Aviles and A. Kouki, "Position-aided mm-wave beam training under nlos conditions," *IEEE Access*, vol. 4, pp. 8703–8714, 2016.
- [15] Z. Wei, Y. Zhao, X. Liu, and Z. Feng, "Doa-If: A location fingerprint positioning algorithm with millimeter-wave," *IEEE Access*, vol. 5, pp. 22678–22688, 2017.
- [16] K. Satyanarayana, M. El-Hajjar, A. A. M. Mourad, and L. Hanzo, "Deep learning aided fingerprint-based beam alignment for mmwave vehicular communication," *IEEE Trans. Veh. Technol.*, vol. 68, no. 11, pp. 10858–10871, Nov 2019.
- [17] V. Va, T. Shimizu, G. Bansal, and R. W. Heath, "Online learning for position-aided millimeter wave beam training," *IEEE Access*, vol. 7, pp. 30507–30526, 2019.
- [18] M. Hashemi, A. Sabharwal, C. E. Koksall, and N. B. Shroff, "Efficient beam alignment in millimeter wave systems using contextual bandits," in *IEEE INFOCOM 2018*, April 2018, pp. 2393–2401.
- [19] M. Cheng, J. Wang, J. Wang, M. Lin, Y. Wu, and H. Zhu, "A fast beam searching scheme in mmwave communications for high-speed trains," in *2019 IEEE ICC*, May 2019, pp. 1–6.
- [20] 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, 2019.
- [21] 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, 2019.
- [22] K. P. Murphy, *Machine learning: a probabilistic perspective*. MIT press, 2012.
- [23] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*. MIT Press, 2006.
- [24] T. Bai, V. Desai, and R. W. Heath, "Millimeter wave cellular channel models for system evaluation," in *2014 IEEE ICNC*. IEEE, 2014, pp. 178–182.
- [25] G. R. Muns, K. V. Mishra, C. B. Guerra, Y. C. Eldar, and K. R. Chowdhury, "Beam alignment and tracking for autonomous vehicular communication using ieee 802.11ad-based radar," in *IEEE INFOCOM WKSHPS*, April 2019, pp. 535–540.
- [26] G. Lindgren, "Stationary stochastic processes. theory and applications," *Texts in Statistical Science*, vol. 66, no. 3, pp. 360–361, 2012.
- [27] D. Russo and B. Van Roy, "Learning to optimize via posterior sampling," *Mathematics of Operations Research*, vol. 39, no. 4, pp. 1221–1243, 2014.
- [28] S. Bubeck and N. Cesa-Bianchi, "Regret analysis of stochastic and nonstochastic multi-armed bandit problems," *Foundations and Trends in Machine Learning*, vol. 5, no. 1, pp. 101–112, 2012.
- [29] B. C. May, N. Korda, A. Lee, and D. S. Leslie, "Optimistic bayesian sampling in contextual-bandit problems," *Journal of Machine Learning Research*, vol. 13, no. 1, pp. 2069–2106, 2012.
- [30] T. Lattimore and C. Szepesvari, *Bandit Algorithms*. Cambridge University Press, 2020.
- [31] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2006.
- [32] J. Snoek, O. Rippel, K. Swersky, R. Kiros, N. Satish, N. Sundaram, M. M. A. Patwary, Prabhat, and R. P. Adams, "Scalable bayesian optimization using deep neural networks," *International Conference on Machine Learning*, 2015.
- [33] C. Riquelme, G. Tucker, and J. Snoek, "Deep bayesian bandits show-down: An empirical comparison of bayesian deep networks for thompson sampling," *International Conference on Learning Representations*, 2018.
- [34] J. Zhang, Y. Huang, Y. Zhou, and X. You, "Beam alignment and tracking for millimeter wave communications via bandit learning," *IEEE Trans. Commun.*, pp. 1–1, 2020.
- [35] R. J. Adler, *The Geometry of Random Fields*. WILEY, 1981.



**Jianjun Zhang** (S'16-M'18) received the M.S. degree from Nanjing University of Aeronautics and Astronautics, Nanjing, China, in 2014, and the Ph.D. degree from Southeast University, Nanjing, China, in 2018. He is currently a Research Fellow of the electrical and electronics engineering with University College London (UCL), U.K. He was the recipient of the Best Paper Award in the IEEE Globecom 2019. His current research interests include machine learning and optimization, intelligent communications, and probability theory and its applications.



**Christos Masouros** (Senior Member, IEEE) received the Diploma degree in Electrical and Computer Engineering from the University of Patras, Greece, in 2004, and MSc by research and PhD in Electrical and Electronic Engineering from the University of Manchester, UK in 2006 and 2009 respectively. In 2008 he was a research intern at Philips Research Labs, UK. Between 2009-2010 he was a Research Associate in the University of Manchester and between 2010-2012 a Research Fellow at Queen's University Belfast. In 2012 he joined

University College London as a Lecturer. He has held a Royal Academy of Engineering Research Fellowship between 2011-2016.

He is currently a Full Professor in the Information and Communications Engineering research group, Dept. Electrical and Electronic Engineering, University College London. His research interests lie in the field of wireless communications and signal processing with particular focus on Green Communications, Large Scale Antenna Systems, Cognitive Radio, interference mitigation techniques for MIMO and multicarrier communications. He was the recipient of the Best Paper Awards in the IEEE GlobeCom 2015 and IEEE WCNC 2019 conferences, and has been recognised as an Exemplary Editor for the IEEE Communications Letters, and as an Exemplary Reviewer for the IEEE Transactions on Communications. He is an Editor for IEEE Transactions on Communications, and IEEE Transactions on Wireless Communications. He has been an Associate Editor for IEEE Communications Letters, and a Guest Editor for IEEE Journal on Selected Topics in Signal Processing issues Exploiting Interference towards Energy Efficient and Secure Wireless Communications and Hybrid Analog / Digital Signal Processing for Hardware-Efficient Large Scale Antenna Arrays. He is currently an elected member of the EURASIP SAT Committee on Signal Processing for Communications and Networking.