Reinforcement Learning for Optimized Beam Training in Multi-Hop Terahertz Communications

Arian Ahmadi and Omid Semiari
Department of Electrical and Computer Engineering, University of Colorado Colorado Springs, Colorado Springs, CO
Email: {aahmadi, osemiari}@uccs.edu

***Abstract*—Communication at terahertz (THz) frequency bands is a promising solution for achieving extremely high data rates in next-generation wireless networks. While the THz communication is conventionally envisioned for short-range wireless applications due to the high atmospheric absorption at THz frequencies, multi-hop directional transmissions can be enabled to extend the communication range. However, to realize multi-hop THz communications, conventional beam training schemes, such as exhaustive search or hierarchical methods with a fixed number of training levels, can lead to a very large time overhead. To address this challenge, in this paper, a novel *hierarchical beam training scheme with dynamic training levels* is proposed to optimize the performance of multi-hop THz links. In fact, an optimization problem is formulated to maximize the overall spectral efficiency of the multi-hop THz link by dynamically and jointly selecting the number of beam training levels across all the constituent single-hop links. To solve this problem in presence of unknown channel state information, noise, and path loss, a new reinforcement learning solution based on the multi-armed bandit (MAB) is developed. Simulation results show the fast convergence of the proposed scheme in presence of random channels and noise. The results also show that the proposed scheme can yield up to 75% performance gain, in terms of spectral efficiency, compared to the conventional hierarchical beam training with a fixed number of training levels.**

I. Introduction

Future wireless networks are expected to support a new breed of wireless technologies that not only require very high data rates, but also mandate very low communications latency [1]. Among these emerging services include, but not limited to, factory automation, autonomous vehicular platoon systems, swarm of unmanned aerial vehicles (UAVs), and user interactions via wireless extended reality applications [2]. Despite their unique service requirements, these applications are similar in a number of key aspects: 1) they mainly rely on direct device-to-device (D2D) or machine-to-machine (M2M) communications among a group of user equipment (UE), 2) the communication network is formed over *multi-hop* D2D or M2M links, and 3) substantial traffic (e.g., sensing data) must be managed within very short (sub-millisecond) time intervals.

These unique characteristics motivate leveraging the large available bandwidth at very high-frequency bands, particularly over the terahertz (THz) frequencies (collectively considered as $0.1-10$ THz) [3], [4]. In fact, compared with the frequency bands considered in the fifth-generation (5G) new radio specifications[1], THz spectrum can offer an order of magnitude larger bandwidth, suitable for managing large sensing information required in autonomous systems. Additionally, deployment of advanced phased arrays (composed of many antenna elements) with very small form-factors is feasible at THz frequencies. This allows UEs to leverage large array processing gains and form highly directional multi-hop links to cope with the large atmospheric absorption at THz frequency bands, achieve extremely high data rates, and extend the communication range to form larger D2D or M2M networks (e.g., UAV swarm or vehicular platoons).

However, one of the key challenges for establishing directional links at high-frequency bands is the lack of full or even partial knowledge of the channel state information (CSI) at the transceivers during the initial access [5]. Therefore, prior to the actual data transmissions, UEs have to follow a process, known as *beam training*, during which the transceivers direct the antenna array gain toward different directions to find the optimal spatial path that maximizes the received power. As UEs move and the propagation environment changes, the beam training process must be repeated to find the best spatial path and maintain high data rates across the network. Moreover, in multi-hop communications, the overall link performance (e.g., data rate) depends on the performance of all the constituent single-hope links [6]. Hence, for multi-hop THz communications, the beam training process must be completed jointly for multiple links at every transmission block, leading to substantial time overhead.

Thus far, substantial work has been done to optimize the beam training process, particularly for communications below 100 GHz [7]–[15]. The authors in [14] propose an exhaustive search beam training, which sequentially tests all possible beam pairs in the angle domain between a base station (BS) and a UE and chooses the precoding/combining codeword that yields maximum received signal power. The main drawback of this approach is that scanning the angular space via sequential search is very time consuming, particularly in THz communications that require very high resolution angular search to achieve the so-called pencil beams. To reduce the beam training time overhead, the hierarchical beam training is developed in [7]–[13], [15]. This technique allows BSs/UEs to scan the angular space with wider beams, and then, narrow down the search space and the beamwidth over multiple training stages. While this approach has been widely adopted to decrease the training time, its performance is highly dependent on the codebook design. In [7], the authors present a fast discovery hierarchical search strategy to decrease the delay of exhaustive search. The authors in [8] propose an analog beamforming strategy using a hierarchical scheme for a single-hop transmission scenario. In [9], an efficient hierarchical codebook is designed by jointly exploiting sub-array and deactivation antenna processing techniques. The work in [15] presents a hierarchical multi-resolution codebook based on hybrid

This research was supported by the U.S. National Science Foundation under Grants CNS-1941348 and CNS-2008646.

[1]In particular, frequency range 1 (sub-6 GHz) and frequency range 2 (sub-100 GHz) millimeter wave (mmWave) bands.

beamforming precoding in a single-UE mmWave system. In [10], the authors introduce a Discrete Fourier Transform (DFT) based multi-level codebook design that yields beam patterns with near-uniform gains at each training level. In [16], the authors propose an online learning algorithm to solve the problem of beam training in mmWave vehicular systems based on a contextual multi-armed bandit (MAB) method. In [11], an online stochastic optimization problem is solved as a unimodal MAB problem to improve beam training in mmWave networks. In fact, the authors utilize the correlation and unimodality properties to decrease the search space and maximize the received energy. The authors in [12] consider a beam-training scheme based on Bayesian MAB to maximize the throughput of the mmWave systems in a single-UE scenario. While the hierarchical beam training schemes developed in [7]–[13], [15], reduce the time overhead compared to the sequential search, they only focus on single-UE, single-hop communications mainly at mmWave frequency bands. As we will show in this paper, the hierarchical beam training schemes can lead to large beam training overhead and performance degradation when applied directly to multi-hop THz links. In fact, the time overhead of existing hierarchical beam training schemes can scale linearly with the number of UEs, which makes them inefficient for multi-hop THz communications.

The main contribution of this paper is a novel hierarchical beam training scheme to reduce the time overhead of the beam training process and enhance the performance for multi-hop THz communication links. To this end, we formulate an optimization problem that aims to maximize the overall spectral efficiency of the multi-hop THz link by finding the optimal number of training levels in hierarchical beam search, jointly for all the constituent single-hop links. In particular, instead of performing the hierarchical search for a pre-defined number of training levels, the proposed scheme dynamically determines the number of training levels for each single-hop link while considering the performance of other links. Hence, the proposed scheme can effectively reduce the beam training overhead and increase the available time for data communication during each transmission block. To solve this problem, we propose a new algorithm that builds on an MAB strategy to efficiently learn the optimal values for the number of search levels during the hierarchical beam training, without requiring prior knowledge about the CSI. The simulation results show that compared with the conventional hierarchical beam training with a fixed number of training levels, the proposed algorithm yields up to 75% performance gain in terms of spectral efficiency.

The rest of the paper is organized as follows. Section II presents the system model. Section III describes the problem formulation based on the proposed hierarchical beam training with dynamic training levels. The proposed algorithm is presented in Section IV. Simulation results are provided in Section V and conclusions are presented in Section VI.

II. System Model

Consider a network of K UEs in a set $\mathcal{K}$ that communicate with one another over a multi-hop THz link. Let $u_1 \leftrightarrow u_2 \leftrightarrow \cdots \leftrightarrow u_K$ denote the multi-hop THz link where $u_k \leftrightarrow u_{k+1}$ represents a single-hop bi-directional THz link between UE u_k to UE u_{k+1}. As an example, Fig. 1 shows a two-hop THz communication $u_1 \leftrightarrow u_2 \leftrightarrow u_3$. We note that more complex structures for the multi-hop network (e.g., mesh or star networks) can be built based on the considered structure, i.e., a connected polytree graph with each node having a maximum degree of two. To form the directional links, each UE u_k is equipped with a uniform linear array composed of N antenna elements.

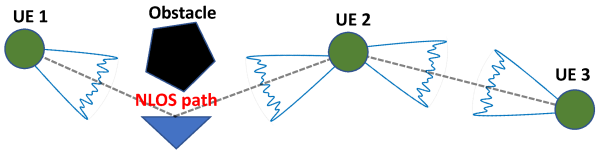

Fig. 1. An example of multi-hop THz communication composed of line-of-sight (LOS) and non-LOS (NLOS) D2D/M2M links.

A. Channel model

Given an azimuth steering direction $\psi \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$, the transceiver's response vector of a UE is given by

$$\boldsymbol{a}(\theta) = \left[1, e^{-j2\pi\frac{d}{\lambda}\theta}, \cdots, e^{-j2\pi\frac{d}{\lambda}(N-1)\theta}\right]^T, \quad (1)$$

where d is the antenna spacing, λ is the wavelength, and $\theta = \sin(\psi)$. For the channel model at high-frequency bands, it is widely accepted to consider an LOS link or an NLOS link with a single spatial path associated with a cluster of scatterers (as shown in Fig. 1) [10], [12]. In fact, for an arbitrary single-hop THz link $u_i \leftrightarrow u_j$, the multiple-input multiple-output (MIMO) channel matrix can be written as

$$\mathbf{H}_{ij} = \beta \boldsymbol{a}(\theta_i)\boldsymbol{a}^H(\theta_j), \quad (2)$$

where θ_i and θ_j are, respectively, the angle-of-departure (AoD) and the angle-of-arrival (AoA) associated with the spatial path for the link between UEs u_i and u_j. In addition, β is the small-scale fading channel gain which is modeled as a zero-mean, complex Gaussian random variable with variance σ_β^2, i.e., $\beta \sim \mathcal{CN}(0, \sigma_\beta^2)$.

B. Beam training and data transmissions

We consider the transmission of data frames, each composed of L time slots. Denoting T_c as the channel coherence time and τ as the duration of each time slot, L is selected such that $L\tau \ll T_c$. At the beginning of each frame, L' time slots are allocated for the beam training between the transmitter and the receiver of a THz link. Hence, $L - L'$ time slots will be assigned for the transmission of data symbols. For an arbitrary link from UE u_i to UE u_j, the received signal over the MIMO channel at a given time slot can be represented as

$$r_j = \sqrt{p_j}\boldsymbol{v}^H\mathbf{H}_{ij}\boldsymbol{w}x + \boldsymbol{v}^H\boldsymbol{n}, \quad (3)$$

where x is the transmitted symbol with $\mathbb{E}\{|x|^2\} = 1$ and p_j is the omni-directional received power (i.e., transmit power after impacted by the path loss) at the receiver j. The additive white Gaussian noise (AWGN) vector $\boldsymbol{n} \in \mathbb{C}^N$ has a zero mean with $\mathbb{E}\{\boldsymbol{n}\boldsymbol{n}^H\} = \sigma_n^2\mathbf{I}_N$. Moreover, $\boldsymbol{w} \in \mathbb{C}^N$ and $\boldsymbol{v} \in \mathbb{C}^N$ represent, respectively, the beamforming and combining vectors. These vectors are selected from a predefined codebook and satisfy $\|\boldsymbol{w}\|^2 = \|\boldsymbol{v}\|^2 = 1$. With this

model, the received signal-to-noise ratio (SNR) is

$$\gamma_{i,j} = \frac{p_j}{\sigma_n^2} |\boldsymbol{v}^H \mathbf{H}_{ij} \boldsymbol{w}|^2. \quad (4)$$

Accordingly, the spectral efficiency of a single-hop link between UEs u_i and u_j is

$$R_{i,j} = [1 - \mathbb{P}(\gamma_{i,j} < \gamma_{\text{th}})] \left(1 - \frac{L'_{i,j}}{L}\right) \log_2(1 + \gamma_{i,j}), \quad (5)$$

where γ_{th} denotes the minimum required SNR and $\mathbb{P}(\gamma_{i,j} < \gamma_{\text{th}})$ represents the outage probability. Next, we describe why the time overhead of beam training in multi-hop THz communications can severely impact the performance and we justify the need for new solutions to optimize the beam training for multi-hop THz links.

III. Multi-Resolution Beam Training With Dynamic Training Levels

Here, we first focus on analyzing the beam training time overhead L' in multi-hop THz networks. As described in Sec. I, the sequential search will result in a significant time overhead to establish bi-directional THz links. Therefore, in this section, we first briefly overview the hierarchical (also known as multi-resolution) search as a widely adopted beam training scheme and analyze its time overhead for multi-hop THz communications. Then, we propose an optimization problem to maximize the performance of multi-hop THz links by effectively reducing the beam training time overhead.

A. Time overhead of hierarchical beam training for multi-hop THz communications

The hierarchical beam training is a technique which allows a UE to start the search using codewords with wide beamwidths, and then, fine-tune the search (i.e., narrow down the angular search space and the beamwidth) at each subsequent level throughout the search process. While different multi-resolution codebooks have been proposed in the literature, here, we build our framework based on the phase-shifted DFT codebook introduced in [10] due to: 1) efficient DFT-based implementation, and 2) near-uniform antenna gain over the beamwidths. As shown in Fig. 2, the hierarchical beam training process requires M training levels to complete the beam training at a transmitter with $\mathcal{W}^{(m)} = \{\boldsymbol{w}_1^{(m)}, \boldsymbol{w}_2^{(m)}, \cdots, \boldsymbol{w}_{q_m}^{(m)}\}$ representing the set of q_m beamforming codewords at the m-th level. Using the phase-shifted DFT codebook design, we can construct each codeword as

$$\boldsymbol{w}_i^{(m)} = \frac{\sqrt{q_m}}{N} \sum_k a_k(\theta_k) e^{j\omega_m k}, \quad (6)$$

where $(i-1)\frac{N}{q_m} + 1 \leq k \leq i\frac{N}{q_m}$ and $\theta_k = -1 + \frac{2k-1}{N}$. In addition, ω_m represents the phase shift added to reduce the antenna gain fluctuations of the codeword over its beamwidth of $\frac{2\pi}{q_m}$. To perform the beam training for a single link, beam search can start at the transmitter while the receiver operates with an omni-directional array gain and provides the index of best codeword to the transmitter over a feedback channel. Then, the transmitter sends training signals using the selected beam training codeword and the beam training can be repeated at the receiver. To implement this process, we use an s_k-way decision tree which allows a UE k to only send s_k training signals at each level of the hierarchical search and complete the beam training at $M_k = \log_{s_k} N$ levels. Considering that only one training signal is sent per time slot, the time overhead of beam training for a single-hop link $u_k \leftrightarrow u_{k+1}$ will be

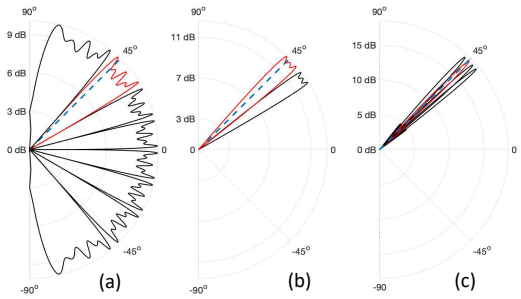

Fig. 2. Example of hierarchical beam training process at the transmitter with $M = 3$, $q_m \in [8, 16, 64]$, and $\omega_m = 2.24$. At each stage, the selected codeword is shown in red. The dashed line represents a 45°AoD.

$$\begin{aligned} L'_{k,k+1} &= s_k M_k + s_{k+1} M_{k+1} \\ &= s_k \log_{s_k} N + s_{k+1} \log_{s_{k+1}} N. \end{aligned} \quad (7)$$

To establish a multi-hop communications with K THz links, we consider a time-division multiple access (TDMA) hierarchical beam training, i.e., the beam training is performed sequentially across the single-hop links. With this in mind, the time overhead increases with the order $O(Ks \log_s N)$, if $s_i = s, \forall i$. For example, for the two-hop link in Fig. 1 with $N = 64$, $s = 4$, the time overhead of the beam training will be $L' = 48$. To better evaluate this time overhead within the total frame duration, let 5 Km/h be the maximum UE velocity. Then, at 240 GHz carrier frequency and with 120 KHz subcarrier spacing, the total number of time slots within the channel coherence time will be $L = 108$. Thus, the beam training overhead for a two-hop THz link will be close to 30% which is significantly large. Here, we note that simultaneous beam training schemes (e.g., in [17]) that are designed for cellular downlink transmissions cannot be applied to distributed D2D/M2M networks considered in this work. That is because such schemes require a central node (e.g., a base station) to separate and group UEs in the spatial domain and manage the interference resulting from concurrent beam trainings. Next, we propose a new beam training approach to reduce this substantial delay overhead in multi-hop THz communications.

B. Problem Formulation

The performance of the multi-hop link $u_1 \leftrightarrow u_2 \leftrightarrow \cdots \leftrightarrow u_K$ depends on the spectral efficiency of the involved single-hop links $u_k \leftrightarrow u_{k+1}$, for $1 \leq k < K$. More generally, with a decode-and-forward scheme, the overall spectral efficiency of the multi-hop THz link between two arbitrary UEs $u_i, u_j \in \mathcal{K}$, with $1 \leq i < K$ and $i < j \leq K$, is

$$R_{i,j} = \left[\prod_{i \leq k < j} \mathbb{P}(\gamma_{k,k+1} \geq \gamma_{\text{th}})\right] \left(1 - \frac{L'_{i,j}}{L}\right) \left(\frac{1}{j-i}\right) \log_2(1 + \gamma_{\text{min}}), \quad (8)$$

where $L'_{i,j} = \sum_k L'_{k,k+1}$, $j-i$ is the number of hops, and $\gamma_{\text{min}} = \min\{\gamma_{k,k+1}\}$ for $i \leq k < j$. In fact, (8) implies that the link with the smallest SNR, γ_{min}, will limit the overall spectral efficiency of the multi-hop communication, irrespective of the beam training and the resulting SNR value at other links [6]. Hence, selecting larger M values at other links (with $\gamma_{k,k+1} > \gamma_{\text{min}}$) will only increase the beam training overhead without increasing the overall spectral efficiency $R_{i,j}$ of the multi-hop THz link. Thus, *we can increase the performance of a multi-hop THz link by optimizing the number of search levels M during the hierarchical beam training*. In fact, given γ_{th}, we must find the optimal value for the number of levels $1 \leq m^* \leq M$ so as to maximize the spectral efficiency of the multi-hop link.

For a single-hop link $u_k \leftrightarrow u_{k+1}$, let $\gamma_{k,k+1}^{(m_k)}$ denote the SNR when beam training is performed up to the m_k-th level at the transceivers. This SNR can be calculated by substituting the selected beamforming and combining vectors at the m_k-th level in (4). With this in mind, we aim to find the optimal vector $\boldsymbol{m} = [m_i, m_{i+1}, \cdots, m_{j-1}]$ for the multi-hop THz link, such that

$$\underset{\boldsymbol{m}}{\text{argmax}} \;\; R_{i,j} \tag{9a}$$

$$\text{s.t.,} \;\; \gamma_{\text{min}} = \min\left\{\gamma_{k,k+1}^{(m_k)}\right\}, \;\; i \leq k < j, \tag{9b}$$

$$L'_{i,j} < \left\lfloor \frac{L}{K-1} \right\rfloor (j-i), \tag{9c}$$

$$m_k \in \{1, 2, \cdots, M_k\}, \;\; i \leq k < j. \tag{9d}$$

The first constraint in (9b) shows that γ_{min} is calculated based on the SNR at the selected search levels. Assuming that the total number of time slots is uniformly allocated to each single-hop link, and let $\lfloor . \rfloor$ denote the floor operation, then, the constraint in (9c) guarantees that the time overhead of beam training is less than the allocated time slots to the THz link with $j-i$ hops. In addition, the feasibility constraint in (9d) ensures that m_k at the transceivers of the $k-i+1$-th single-hop link does not exceed the number of training levels in the conventional multi-resolution beam training. Next, we develop a new approach to solve the proposed problem for hierarchical beam training with *dynamic training levels*.

IV. Proposed Hierarchical Beam Training Algorithm for Multi-Hop THz Communications

Clearly, the objective function in (9a) is not a monotonic function of $\boldsymbol{m}$, since $\log_2(1+\gamma_{\text{min}})$ can be an increasing function of m_k parameters, while the pre-log factor, $1 - L'_{i,j}/L$, is a decreasing function m_k variables in $\boldsymbol{m}$. To solve the proposed problem in (9a)-(9d), we note that it is very challenging to derive the outage probability as a function of $\boldsymbol{m}$. Moreover, the CSI of the MIMO channel for each link and γ_{min} are not known prior to the beam training phase. Hence, it is not feasible to solve the proposed problem in (9a)-(9d) via standard optimization techniques. To this end, we develop a new beam training approach, based on reinforcement learning, to find the optimal $\boldsymbol{m}$ for a multi-hop THz link while considering the joint performance of its constituent single-hop links.

In fact, we can formulate the optimization problem in (9a)-(9d) as an MAB problem, in which the transceivers of the multi-hop link (acting as the agents) explore different choices for a vector $\boldsymbol{l}$ (analogous to an arm in an MAB problem) with the k-th element being $l_k = M_k - m_k = \log_{s_k} N - m_k$ for $i \leq k < j$. Here, the integer variable l_k ($0 \leq l_k < M_k$) represents the number of reduced search levels for beam training at the transmitter and receiver of the link k. After playing an arm $\boldsymbol{l}$ from a set of all possible arms $\mathcal{L}$, the UEs of the multi-hop link will receive a random reward $P_{i,j}(\boldsymbol{l})$ which is equal to the spectral efficiency in (8). To determine the size of the set $\mathcal{L}$, we note that each l_k can take M_k integer values from 0 to $M_k - 1$. Therefore, for the multi-hop link $u_1 \leftrightarrow u_2 \leftrightarrow \cdots \leftrightarrow u_K$, the total number of arms will be equal to $\Pi_{k=1}^{K-1} M_k$.

TABLE I
Proposed Hierarchical Beam Training Algorithm with Dynamic Training Levels

Inputs: $\mathcal{K}$, γ_{th}, ω_m, ε_0, s_k, $t = 0$.
while *($t \leq T$)* **do**
Step 1: Optimize the number of training levels using the epsilon-decay strategy:
a. With probability $1 - \varepsilon_t$, select the arm with the current maximum average reward. Otherwise, select an arm $\boldsymbol{l}^t \in \mathcal{L}$ randomly. Then, increase t to $t+1$.
b. Update the value of ε_t using $\varepsilon_t = \varepsilon_{t-1}(1000/(1000+t))$.
Step 2:
a. Using the selected arm, follow the hierarchical beam training described in Section III.
b. Calculate the reward from (8) and update the average reward for the selected arm.
end
Output: The arm with maximum average reward.

The key advantage of the MAB-based solution is that it enables the transceivers across the multi-hop THz link to jointly find the optimal solution for (9a)-(9d), in presence of stochastic noise and channel variations. The transceivers of the multi-hop link can try only one arm $\boldsymbol{l} \in \mathcal{L}$ at each trial (i.e., block transmission of L time slots). Within this MAB framework, we define the regret $\zeta(T)$ after T trial as

$$\zeta(T) = \sum_{t=1}^{T} P_{i,j}(\boldsymbol{l}^*) - P_{i,j}(\boldsymbol{l}^t), \tag{10}$$

where $\boldsymbol{l}^*$ is the optimal arm with elements $l_k^* = M_k - m_k^*$ where $m_k, i \leq k < j$ is the solution of the proposed problem in (9a)-(9d) and $\boldsymbol{l}^t$ is the played arm at round t with an associated reward $P_{i,j}(\boldsymbol{l}^t)$. The objective is to find a strategy that selects $\boldsymbol{l}^t$, for $1 \leq t \leq T$, such that $\lim_{T\to\infty} \zeta(T) = 0$. Clearly, such strategy will converge to the solution of the proposed problem in (9a)-(9d).

The proposed algorithm is summarized in Table I which builds on the epsilon-decay strategy to solve the MAB problem. The reason for employing the epsilon-decay strategy is that it can properly maintain the tradeoff between exploration versus exploitation during the learning process, particularly if the size of the set $\mathcal{L}$ is not too large. With this in mind, for a given set of input parameters, the proposed algorithm follows a two-step process during each trial (i.e., a transmission block of L time slots). In Step 1 of an arbitrary trial t, the transceivers of the multi-hop link select an arm $\boldsymbol{l}^t \in \mathcal{L}$ based on the epsilon-decay strategy. That is, the algorithm chooses a random arm with probability ε_t or selects the arm with the highest current average reward with probability $1-\varepsilon_t$. Once an arm is selected, the transceivers follow the hierarchical beam search according to the selected arm. That is, the transceivers of the k-

TABLE II
SIMULATION PARAMETERS

Notation	Parameter	Value
f_c	Carrier frequency	240 GHz
N	Number of antennas	64
s_k	Number of training signals	4
ω_m	Phase shift parameter of the DFT-based codebook	2.24 rad/s
v	UE maximum speed	5 km/h
σ_β	Channel gain standard deviation	1
–	Subcarrier spacing	120 kHz [18]
γ_{th}	Minimum required SNR	-50 dB
N_0	Noise power spectral density	-204 dBm/Hz [19]
–	Path loss exponent	2.02 dB [18]
B	Total system bandwidth	4 GHz [19]

th single-hop link will follow the beam training process explained in Sec. III for up to $m_k = M_k - l_k$ training levels. After receiving the instantaneous reward in (8) for the t-th trial, the average reward for the selected arm will be updated and the process is repeated until up to T arms are played. The output of the algorithm will be the arm l^* with the maximum average reward.

V. SIMULATION RESULTS

In this section, we present the simulation results and show the performance of the proposed algorithm in terms of its convergence, the statistics of the achievable spectral efficiency, and the probability of miss detection at the receivers of the multi-hop THz link. For simulations, we consider three UEs communicating with one another over a two-hop THz link, as shown in Fig. 1. The distances between UEs 1 and 2 and UEs 2 and 3 are, respectively, 30 m and 5 m. The received SNR in (4) is calculated at the output of the matched filter where the length of the filter is calculated based on sampling the training signal at the Nyquist rate. Both links have the same transmit SNR, ranging from 20 dB to 60 dB. Simulation parameters are summarized in Table II. We compare the performance of the proposed beam training with dynamic training levels (labeled as "Hierarchical, Dynamic") with two other baseline schemes: 1) The conventional hierarchical beam training described in Sec. III-A (labeled as "Hierarchical, Fixed"), and 2) The hierarchical beam training with a random number of training levels (labeled as "Hierarchical, Random"). The performance was evaluated by averaging the results over sufficiently large Monte Carlo runs.

Figure. 3 shows the average regret resulting from the proposed MAB-based algorithm versus the number of trials for different values of the transmit SNR. Here, the average regret is computed by averaging the regret in (12) over large independent runs. From Fig. 3, we observe that the regret decreases rapidly, showing the fast convergence of the proposed learning approach in presence of channel fading, random AoA/AoD, and the receiver noise. The results show that the proposed beam training algorithm successfully converges to the optimal solution within a reasonably small number of trials. As an example, for 40 dB transmit SNR, the average regret will be less that 10^{-5} after 100 trials.

In Fig. 4, we compare the average spectral efficiency of the two-hop THz link versus the transmit SNR for the proposed approach and the two baseline schemes. Clearly, the spectral efficiency increases with higher transmit SNR values. The results in Fig. 4 also show that the proposed algorithm outperforms the other two schemes, with up to

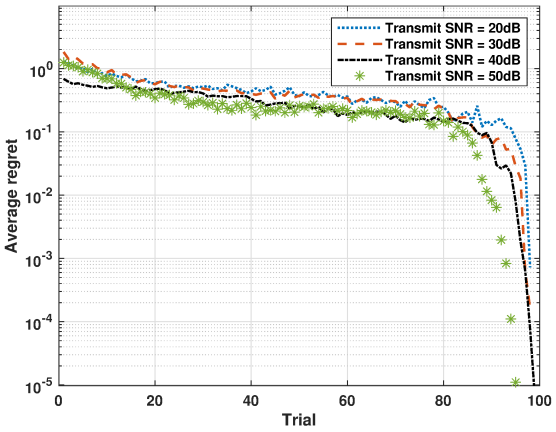

Fig. 3. Average regret versus the number of trials.

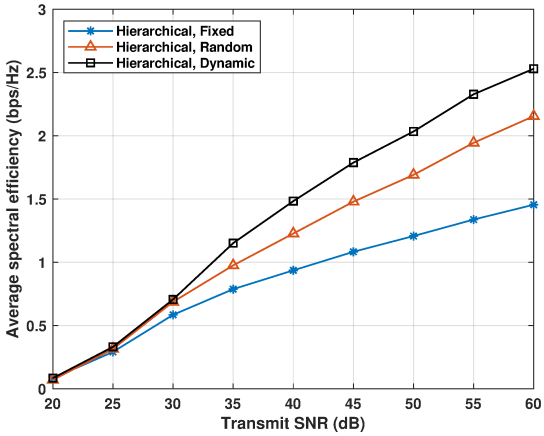

Fig. 4. Average spectral efficiency of the two-hop THz link versus the transmit SNR.

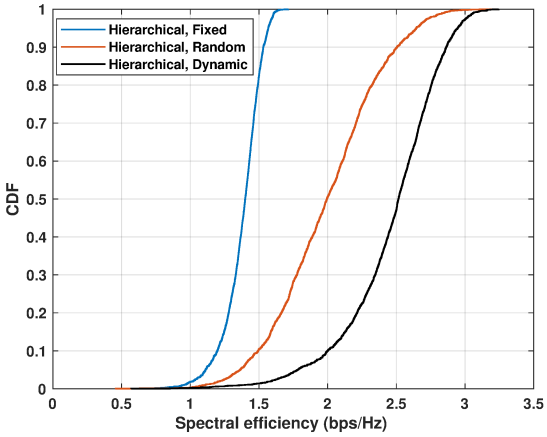

Fig. 5. The CDF of spectral efficiency of the two-hop THz link at 50 dB transmit SNR.

17% and 75% performance gains compared to, respectively, the "Hierarchical, Random" and "Hierarchical, Fixed" approaches at 60 dB transmit SNR.

Figure. 5 presents the cumulative distribution function (CDF) of the spectral efficiency of the two-hop THz link resulting from the proposed algorithm and the two baseline beam training schemes at 50 dB transmit SNR. The figure indicates that although the "Hierarchical, Random" algorithm is more efficient than the "Hierarchical, Fixed" scheme, its performance is not comparable to the proposed approach. In fact, the figure shows that the proposed algorithm achieves a spectral efficiency greater than 2 bps/Hz

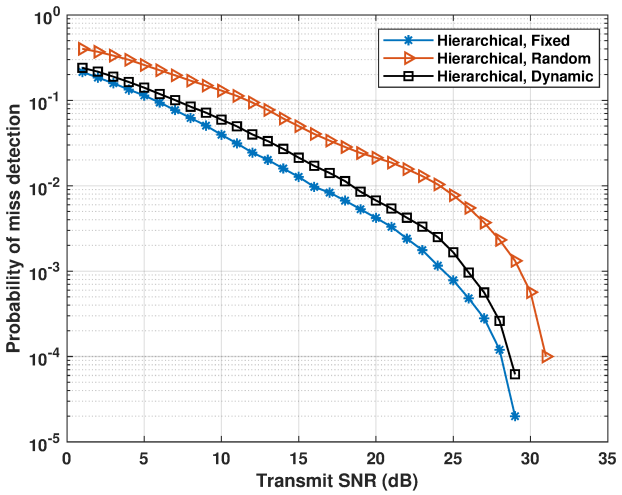

Fig. 6. Probability of miss detection versus transmit SNR.

with about 0.9 probability. However, this probability is only 0.5 for the "Hierarchical, Random" scheme while the "Hierarchical, Fixed" cannot achieve the 2 bps/Hz spectral efficiency.

While the proposed scheme can effectively optimize the performance of the multi-hop THz link (as shown in Figs. 3-5), reducing the number of training levels can impact the detection of training signals from noise. In fact, as the proposed scheme reduces the number of training levels, the effective antenna gain can be reduced and the receivers' detectors may not be able to detect the actual training signal from noise particularly at low transmit SNR values. To study this effect, Fig. 6 compares the probability of miss detection of the proposed algorithm with the two other baseline schemes. For the signal detection, we use the Neyman-Pearson detector with a fixed probability of false alarm (P_{FA}) of 0.01. Based on the results in Fig. 6, the probability of miss detection decreases as we increase the transmit SNR. Moreover, Fig. 6 shows that the performance degradation for the proposed scheme is negligible compared to the "Hierarchical, Fixed" scheme. Comparing the results of Figs. 4-6, we observe that the proposed scheme can significantly improve the spectral efficiency of multi-hop THz communications without any major impact on the detection of training signals.

VI. Conclusions

In this paper, we have proposed a novel beam training approach, based on reinforcement learning to optimize the performance of multi-hop THz communication links. First, we have shown the substantial time overhead of conventional hierarchical beam training schemes when applied to multi-hop THz communications. Then, to address this challenge, we have introduced a new hierarchical beam training scheme with dynamic training levels to effectively reduce the time overhead of the beam training process and maximize the multi-hop link's performance. To find the optimal number of training levels across the multi-hop link, we have formulated an optimization problem that maximizes the spectral efficiency, while considering the beam training time constraints. To solve the problem with no prior CSI knowledge of the links during the beam training phase, we proposed an MAB-based algorithm that can effectively find the optimal training levels with reasonably fast convergence. The simulation results have shown that the proposed approach can yield up to 75% and 17% performance gains in spectral efficiency, compared to, respectively, the hierarchical beam training with a fixed and random number of training levels.

References

[1] O. Semiari, W. Saad, M. Bennis, and M. Debbah, "Integrated millimeter wave and sub-6 GHz wireless networks: A roadmap for joint mobile broadband and ultra-reliable low-latency communications," *IEEE Wireless Communications*, vol. 26, no. 2, pp. 109–115, 2019.

[2] T. Rappaport, Y. Xing, O. Kanhere, S. Ju, A. Madanayake, S. Mandal, A. Alkhateeb, and G. Trichopoulos, "Wireless communications and applications above 100 GHz: Opportunities and challenges for 6G and beyond," *IEEE Access*, vol. 7, pp. 78729–78757, 2019.

[3] N. Rajatheva, I. Atzeni, et al., "White paper on broadband connectivity in 6G," *arXiv preprint arXiv:2004.14247*, 2020.

[4] R. Barazideh, O. Semiari, S. Niknam, and B. Natarajan, "Reinforcement learning for mitigating intermittent interference in terahertz communication networks," in *2020 IEEE International Conference on Communications Workshops (ICC Workshops)*, 2020, pp. 1–6.

[5] C. Barati, S. Hosseini, M. Mezzavilla, T. Korakis, S. Panwar, S. Rangan, and M. Zorzi, "Initial access in millimeter wave cellular systems," *IEEE Transactions on Wireless Communications*, vol. 15, no. 12, pp. 7926–7940, 2016.

[6] M. Sikora, J. N. Laneman, M. Haenggi, D. J. Costello, and T. E. Fuja, "Bandwidth- and power-efficient routing in linear wireless networks," *IEEE Transactions on Information Theory*, vol. 52, no. 6, pp. 2624–2633, 2006.

[7] V. Desai, L. Krzymien, P. Sartori, W. Xiao, A. Soong, and A. Alkhateeb, "Initial beamforming for mmwave communications," in *2014 48th Asilomar Conference on Signals, Systems and Computers*. IEEE, 2014, pp. 1926–1930.

[8] J. Wang, Z. Lan, et al., "Beam codebook based beamforming protocol for multi-Gbps millimeter-wave WPAN systems," *IEEE Journal on Selected Areas in Communications*, vol. 27, no. 8, pp. 1390–1399, 2009.

[9] Z. Xiao, T. He, P. Xia, and X. Xia, "Hierarchical codebook design for beamforming training in millimeter-wave communication," *IEEE Transactions on Wireless Communications*, vol. 15, no. 5, pp. 3380–3392, 2016.

[10] S. Noh, M. Zoltowski, and D. Love, "Multi-resolution codebook and adaptive beamforming sequence design for millimeter wave beam alignment," *IEEE Transactions on Wireless Communications*, vol. 16, no. 9, pp. 5689–5701, 2017.

[11] M. Hashemi, A. Sabharwal, C. Koksal, and N. Shroff, "Efficient beam alignment in millimeter wave systems using contextual bandits," in *IEEE INFOCOM 2018-IEEE Conference on Computer Communications*. IEEE, 2018, pp. 2393–2401.

[12] M. Hussain and N. Michelusi, "Second-best beam-alignment via bayesian multi-armed bandits," in *2019 IEEE Global Communications Conference (GLOBECOM)*. IEEE, 2019, pp. 1–6.

[13] 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 transactions on communications*, vol. 61, no. 10, pp. 4391–4403, 2013.

[14] C. Jeong, J. Park, and H. Yu, "Random access in millimeter-wave beamforming cellular networks: issues and approaches," *IEEE Communications Magazine*, vol. 53, no. 1, pp. 180–185, 2015.

[15] A. Alkhateeb, O. El Ayach, G. Leus, and R. Heath, "Channel estimation and hybrid precoding for millimeter wave cellular systems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 5, pp. 831–846, 2014.

[16] A. Asadi, S. Müller, G. Sim, A. Klein, and M. Hollick, "Fml: Fast machine learning for 5G mmwave vehicular communications," in *IEEE INFOCOM 2018-IEEE Conference on Computer Communications*. IEEE, 2018, pp. 1961–1969.

[17] R. Zhang, H. Zhang, W. Xu, and C. Zhao, "A codebook based simultaneous beam training for mmwave multi-user mimo systems with split structures," in *2018 IEEE Global Communications Conference (GLOBECOM)*, 2018, pp. 1–6.

[18] Y. Xing and T. Rappaport, "Propagation measurement system and approach at 140 GHz-moving to 6G and above 100 GHz," in *2018 IEEE Global Communications Conference (GLOBECOM)*. IEEE, 2018, pp. 1–6.

[19] A. Ekti, A. Boyaci, A. Alparslan, İ. Ünal, S. Yarkan, A. Görçin, H. Arslan, and M. Uysal, "Statistical modeling of propagation channels for terahertz band," in *2017 IEEE Conference on Standards for Communications and Networking (CSCN)*. IEEE, 2017, pp. 275–280.