DiVA
http://www.diva-portal.org

This is the published version of a paper presented at *IEEE Wireless Communications and Networking Conference, 26–29 March 2023, Glasgow, Scotland, UK*.

Citation for the original published paper:

Khosravi, S., S. Ghadikolaei, H., Zander, J., Petrova, M. (2023)
Reinforcement Learning-based Joint Handover and Beam Tracking in Millimeter-wave Networks
In:

N.B. When citing this work, cite the original published paper.

Permanent link to this version:
http://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-326535

# Reinforcement Learning-based Joint Handover and Beam Tracking in Millimeter-wave Networks

Sara Khosravi*, Hossein S. Ghadikolaei‡, Jens Zander*, and Marina Petrova *†

*School of EECS, KTH Royal Institute of the Technology, Stockholm, Sweden,
† Mobile Communications and Computing, RWTH Aachen University, Germany, ‡ Ericsson Research, Sweden
Email: {sarakhos, jenz, petrovam} @kth.se, hossein.shokri.ghadikolaei@ericsson.com

***Abstract*—In this paper, we develop an algorithm for joint handover and beam tracking in millimeter-wave (mmWave) networks. The aim is to provide a reliable connection in terms of the achieved throughput along the trajectory of the mobile user while preventing frequent handovers. We model the association problem as an optimization problem and propose a reinforcement learning-based solution. Our approach learns whether and when beam tracking and handover should be performed and chooses the target base stations. In the case of beam tracking, we propose a tracking algorithm based on measuring a small spatial neighbourhood of the optimal beams in the previous time slot. Simulation results in an outdoor environment show the superior performance of our proposed solution in achievable throughput and the number of handovers needed in comparison to a multi-connectivity baseline and a learning-based handover baseline.**

***Index Terms*—Millimeter-wave, user association, beam tracking, handover, reinforcement learning.**

## I. Introduction

Millimeter-wave (mmWave) is a key radio access technology for beyond 5G communication systems, offering ultra-high data rates due to a large amount of free spectrum [1]. However, due to the fewer scattering paths and significant penetration loss, mmWave links are vulnerable to static or dynamic obstacles. To overcome such severe loss, both base station (BS) and user equipment (UE) may need directional communication using a large number of antennas, which may result in frequent misalignment of beams due to mobility and blockage. Hence, finding and maintaining the optimal beam directions (beam alignment) is necessary. The lengthy period to achieve the beam alignment (hundreds of milliseconds to seconds [2]) results in a high cell search time or BS discovery time in mmWave systems. As reported in [3], the BS discovery time which is the time required to search the target BS when the handover command is received by the UE is about 200 ms. Moreover, to improve the capacity and coverage the density of the BSs is usually high in mmWave systems [1]. Hence, conventional handover methods based on instantaneous received signal power can cause unnecessarily frequent handovers and a ping-pong effect. This leads to a severe drop in service reliability. Therefore, fast BS discovery (finding target BS in the handover process), and efficient handover execution techniques, will be required to use the full promise of mmWave cellular networks.

The spatial mmWave channel can be approximated by a few dominant paths, where each path can be defined with its angle of departure (AoD), angle of arrival (AoA) and gain [4]. Hence, one can only estimate these path parameters instead of a large dimensional channel matrix [5], [6]. The process of identifying the dominant paths is called beam training. However, due to the dynamic environment, frequent beam training may cause high overhead[1]. Temporal correlation of spatial mmWave channel can be employed to accelerate the beam training process by tracking the variation of the dominant path directions [6].

### A. Related Work

To address the link failure and throughput degradation in a dynamic environment, the multi-connectivity technique has been vastly analyzed in literature [7], [8]. In this technique, the UE keeps its connection to multiple BSs (either at mmWave band or sub-6 GHz band). However, power consumption, synchronization and the need for frequent tracking are the main challenges. In the 3GPP standard (release 16) two handover techniques are introduced to improve the link robustness during mobility: dual active protocol stack (DAPS), and conditional handover (CHO) [9]. In the DAPS, the connection to the current serving BS is maintained until the connection to the target BS is fully established. In the CHO, the UE is configured with multiple target BSs. During the handover, the UE can select one of the configured BSs as the target BS during the RRC reconfiguration message. Although CHO can decrease the handover failure probability, it may increase the handover latency if the UE asks for multiple handovers during a single RRC reconfiguration [7].

Applying machine learning as the main decision-maker tool to make the optimal handover decision and choose the target BS has been also studied in the literature [10], [11]. The authors in [10] proposed a reinforcement learning (RL) based handover policy to reduce the number of handovers while keeping the quality of service in heterogeneous networks. In [11] an intelligent handover method based on choosing the backup solution for each serving link to maximize the aggregate rate along a trajectory has been proposed.

[1]Overhead depends on the training time compared with the changes in the environment.

In terms of beam tracking, authors in [12] applied the correlation of spatial mmWave channel in adjacent locations and proposed the beam steering method based on searching over a small angular space in the vicinity of the previously known valid beams. The authors in [6] applied machine learning to the tracking procedure to extract useful information from the history of AoD tracking.

All the aforementioned works only take handover or beam tracking issues into account. Additionally, they do not study the impact of selecting beam tracking and handover on the achieved throughput of the UE along its trajectory and instead focus on the achieved rate as the primary performance metric.

### B. Our Contributions

In this paper, we develop a novel joint handover and beam tracking algorithm in a mmWave network under mobility. The algorithm aims to associate the UEs to BSs that maximize the sum achieved throughput along the trajectory and ensure the achieved throughput in each location of the trajectory is higher than a pre-defined threshold. The user association process is defined as the process of determining whether a user is associated with a particular BS before data transmissions commence. In the case of handover, the UE is associated with a new BS, whereas in the case of beam tracking, the UE remains associated with the serving BS from the previous time slot. The main contributions of our paper are summarized as below:

- *System Modeling*: We model the user association problem as a non-convex optimization problem. Unlike the existing works in the literature, we consider achieved throughput as the main performance metric to measure the effect of handover or beam tracking on the UEs' quality of service.
- *Learning-based Solution*: The objective function in our proposed user association problem highly depends on the user association mechanism. We utilize the reinforcement learning (RL) algorithm to approximate the solution to this problem. The aim is to decide whether to run a beam tracking algorithm or a handover algorithm.
- *Joint Handover and Beam Tracking Algorithm*: In the case of a handover decision, the target BS will be recognized as the output of the RL algorithm. In the case of beam tracking, the search space will be defined based on our proposed tracking algorithm by searching the directions in the small spatial neighbourhood of the previously selected optimal directions.
- *Empirical Evaluation*: We apply ray tracing with a real building data map as the input. The results show the effectiveness of our proposed method in achieving throughput along trajectories and decreasing the number of handovers.

The rest of the paper is organized as follows. We introduce the system model and problem formulation in Section II. In Section III, we propose our method. We present the numerical results in Section IV and, conclude our work in Section V.

*Notations:* Throughout the paper, vectors and scalars are shown by bold lower-case ($\mathbf{x}$) and non-bold ($x$) letters, respectively. The conjugate transpose of a vector $\mathbf{x}$ is represented by $\mathbf{x}^H$. We define set $[M] := \{1, 2, .., M\}$ for any integer $M$. The indicator function $1\{\cdot\}$ equals to one if the constraint inside $\{\cdot\}$ is satisfied.

## II. System Model and Problem Formulation

In this section, first, we introduce the mmWave channel model. Then, we present the user association problem formulation.

We consider a downlink communication with $|\mathcal{B}|$ mmWave BSs, where each is equipped with $N_{\text{BS}}$ antennas, communicating with a single antenna mobile UE. We consider analog beamforming with a single RF chain. We assume all BSs allocate equal resources to their serving UEs. The channel between BS $j \in \mathcal{B}$ and its serving UE during time slot $i$ is [13]:

$$\mathbf{h}_j = \sum_{\ell=1}^{L} h_\ell \mathbf{a}^H(\phi_\ell, \theta_\ell), \tag{1}$$

where $L$ is the number of available paths. Each path $\ell$ has complex gain $h_\ell$ (include path-loss) and horizontal $\phi_\ell$ and vertical $\theta_\ell$, AoD. Due to the notation simplicity, we drop the index $j$ and $i$ from the channel parameters. The array response vector is $\mathbf{a}(.)$ where its exact expression depends on the array geometry and possible hardware impairments. The signal-to-noise ratio (SNR) in time slot $i$ is

$$\text{SNR}_j^{(i)} = \frac{p|\mathbf{h}_j^H \mathbf{f}_j|^2}{\sigma^2}, \tag{2}$$

where $\sigma^2$ is the noise power, $p$ is the transmit power, $\mathbf{f}_j \in \mathcal{C}^{N_{\text{BS}}}$ is the beamforming vector of BS $j$.

We define variable $x_j^{(i)} \in \{0, 1\}$ for $j \in \mathcal{B}$ as an association indicator in time slot $i$, where is equal 1 if UE is associated to the BS $j$ and 0 otherwise. Hence, the achieved rate per second per hertz in time slot $i$ is

$$\text{R}^{(i)} = x_{j_S}^{(i)} \log_2(1 + \text{SNR}_{j_S}^{(i)}) = \sum_{j \in \mathcal{B}} x_j^{(i)} \log_2(1 + \text{SNR}_j^{(i)}),$$

where $j_S$ is the index of the serving BS of the UE during time slot $i$. Here, we assume each UE is served by only one BS.

We define the achievable throughput per hertz of the UE by multiplying its rate by the data transmission time as

$$\Gamma^{(i)} = (1 - \frac{\tau_b^{(i)}}{\tau_c})\text{R}^{(i)}, \tag{3}$$

where, $\tau_b^{(i)}$ is the beam training duration which may have a different value in each time slot $i$, and $\tau_c$ is the duration of the time slot that is a fixed value for all time slots, see Fig. 1.

### A. Beam Training and Beam Tracking

As depicted in Fig. 1a, when the UE is connected to a BS $j \in \mathcal{B}$, initial beam training is performed by sending pilots over all combination of the beam directions in the codebook during $\tau_b$. Based. on the UE's feedback of the received signal strength (or estimated SNR), the best beam pair directions are selected. Then, the BS and the UE would use this

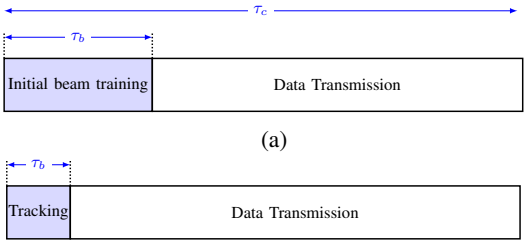


Fig. 1: $\tau_c$ is the time slot duration. $\tau_b$ is (a) the initial beam training duration when the UE is associated with the new BS (handover case), (b) the beam tacking duration when the serving BS is the same for the consecutive slots.

direction $(\phi_{\ell^\star}, \theta_{\ell^\star})$ during the data transmission phase. The beamforming vector, $\mathbf{f}$ is chosen to maximize the achievable rate of the UE. Due to the monotonicity of the logarithm function, this is equivalent to maximising the SNR term in (2). Hence

$$\mathbf{f}_j^* = \underset{\mathbf{f}_j \in \mathcal{F}}{\arg\max} \quad |\mathbf{h}_j^H \mathbf{f}_j|^2 \tag{4}$$

where $\mathcal{F}$ is the beamforming codebook that contains all the feasible beamforming vectors. The n-th element of the codebook $\mathcal{F}$ is defined as $\mathbf{f}(n) = \mathbf{a}(\phi_n, \theta_n)$, where $(\phi_n, \theta_n)$ are steering angles and $\mathbf{a}(.)$ is the array response vector.

When the BS continues serving the same UE in a consecutive time slot, only searching the neighbouring beam directions of the main directions can be sufficient to maintain the link quality. This process is called beam tracking. As shown in Fig. 1b, the duration of $\tau_b$ is much smaller than the initial beam training duration.

### B. Problem Formulation

The UE association depends on the channel quality between the BS and the UE. Due to UE mobility or temporary blockage, the channel quality changes and consequently the UE association. Based on the UEs' velocity, we determine how quickly the channel quality can change and predict the time at which the current UE association needs to be updated. We define $T_{\text{A}}$ seconds as the frequency of updating the association. Hence, we need to make the decision every $T_{\text{A}}$ whether to run the handover execution or beam tracking procedure if SNR is lower than the pre-defined SNR threshold ($\text{SNR}_{\text{thr}}$). Note that we can have an on-demand reactive handover at any time slot if the link toward the serving BS fails abruptly. However, with a proper choice of $T_A$, the frequency of those reactive events could be very small. We define the duration of the trajectory as $M$ and consider the discrete time index $i$ to describe the association update at each interval.

The goal is to maximize the aggregate throughput of the UE along the trajectory while ensuring the achieved throughput in each time slot $i$ is higher than a predefined threshold. To this end, we define functions $F_1$ and $F_2$ as

- $F_1$ is the averaged throughput along the trajectory as

$$F_1 = \sum_{i=1}^{M} \mathbb{E}\left[\Gamma^{(i)}\right],$$

  where the expectation is with respect to the randomness of channel fading and the blockage, $M$ is the duration of the trajectory, and $\Gamma^{(i)}$ is defined in (3).
- $F_2$ is the expected number of time slots whose throughput is lower than the threshold ($\Gamma_{\text{thr}}$).

$$F_2 = \mathbb{E}\left[\sum_{i=1}^{M} 1\left\{\Gamma^{(i)} \leq \Gamma_{\text{thr}}\right\}\right] = \sum_{i=1}^{M} \Pr\left\{\Gamma^{(i)} \leq \Gamma_{\text{thr}}\right\}.$$

We formulate the user association at time slot $i \in [M]$ as an optimization problem which involves finding the $x_j^{(i)}$ corresponding to the association indicator as

$$\max_{\{x_j^{(i)}\}_{i,j}} \quad F_1 - \lambda F_2 \tag{5a}$$

$$\text{s.t.} \quad \sum_{j \in \mathcal{B}} x_j^{(i)} = 1, \forall, i \in [M] \tag{5b}$$

$$x_j^{(i)} \in \{0, 1\}, \quad \forall j \in \mathcal{B}, i \in [M] \tag{5c}$$

where $\lambda$ is a large constant controlling the importance of $F_2$. Constraint (5b) guarantees that each UE is served by one BS.

The optimization problem (5) is nonlinear. Solving this optimization problem requires estimating the expectation value in $F_1$ and $F_2$ which requires running many realizations. Moreover, the impact of choosing the $x_j^{(i)}$ (the target BSs in the handover case or choosing beam tracking procedure) propagates in time and can affect the UEs' performance in the next time slots. Therefore, we need to consider the long-term benefits of selecting association indicators besides their immediate effects on the UEs' performance. Furthermore, In order to select the target BSs, we need to model or predict the UEs' performance in the next time slots, which can add more complexity to the network due to the mobility of the UE and obstacles in mmWave networks. These motivate us to utilize the RL to approximate the solution of (5).

## III. Proposed Method

We transform the problem (5) to an RL problem in which the objective function is turned into a reward function, and the constraints are transformed into the feasible state and action spaces. In the following, first, we start with defining the Markov decision process, and then we will describe our joint handover and beam tracking algorithm.

### A. Markov Decision Process Formulation

RL problems are formulated based on the idea of the Markov decision process (MDP), which is the agent's interaction with different states of the environment to maximize the expected long-term reward. The agent is the main decision-maker who can sit on the edge cloud. All BSs are connected to the agent. Now, we define different elements of an MDP.

*1) State Space:* The state space describes the environment by which the agent is interacting through different actions. We define the state at time slot $i$ as $s^{(i)} = (\ell^{(i)}), j_S^{(i)}, \text{SNR}^{(i)}, I^{(i)}) \in \mathcal{S}$, where $\ell^{(i)}$ is the location index of the UE along the trajectory [2], $j_S^{(i)}$ is the index of the serving BS, $\text{SNR}^{(i)}$ is the SNR value of the UE with serving BS $j_S^{(i)}$ in time slot $i$. $I^{(i)} \in \{0, 1\}$ is the beam tracking activation indicator. $I^{(i)} = 1$ means the $i$-th time slot is the tracking slot for the UE.

*2) Action Space:* The action space includes all possible actions that can be taken by the agent. The action can change the state of the environment from the current state to the target state. In our problem, $a^{(i)} \in \mathcal{A} = \{0, 1, 2, ..., [|\mathcal{B}|]\}$ is the decision regarding beam tracking ($a^{(i)} = 0$) or choosing the index of new serving BS in the case of handover decision ($a^{(i)} \in [|\mathcal{B}|]$). In other words, if $a^{(i)} \neq 0$ means the handover decision is made and the value of $a^{(i)}$ shows the target BS. Hence, the action is to specify a serving BS for the UE along its trajectory.

*3) Policy:* A policy $\pi(.)$ maps the state of the environment to the action of the agent. In our case, $\pi$ is a function from $\mathcal{S}$ to $\mathcal{A}$, i.e., $\pi : \mathcal{S} \to \{0, 1, ..., [|\mathcal{B}|]\}$

*4) Rewards:* The agent obtains the reward after taking an action $a^{(i)}$ when current state is $s^{(i)}$ and moves to next state $s^{(i+1)}$. Here we define reward $r(s^{(i)}, a^{(i)}, s^{(i+1)})$ as

$$r(s^{(i)}, a^{(i)}, s^{(i+1)}) = \Gamma^{(i)} - \lambda 1\left\{\Gamma^{(i)} \leq \Gamma_{\text{thr}}\right\}, \quad (6)$$

where $\Gamma^{(i)}$ is defined in (3).

*5) State-action value:* The function $Q_\pi(s, a)$ is the long-term reward and is defined as the expected summation of discounted reward in the future for the action $a \in \mathcal{A}$ that agent takes in state $s$ under policy $\pi$. The RL algorithm aims to choose the optimal policy $\pi^\star$ in each state $s$ that maximizes the $Q_\pi(s, a)$. With discount factor $\eta \in [0, 1]$, we have

$$Q_\pi(s, a) = \mathbb{E}\left\{\sum_i \eta^i r(s^{(i)}, s^{(i)}, s^{(i+1)})\right\},$$

where the expectation is over the transition probabilities. In our problem, transition probabilities model the SNR variations due to the randomness of the channel fading and blockage. We assume mobility information including the UEs' current location and its trajectory is known[3]. Therefore, the transition to the next location is deterministic.

The optimal policy in state $s \in \mathcal{S}$ is found by

$$\pi^\star(s) = \arg\max_{a \in \mathcal{A}} Q_\pi(s, a). \quad (7)$$

Due to the continuous and large number of state spaces, we apply deep Q-learning (DQL) [14] to solve (7). In DQL, the state-action value function is estimated by the deep neural network function approximators.

[2]Note that, we discretize the location of the UE along the trajectory. Hence, every location dimension $(x, y)$ a trajectory with length $M$ is mapped to a location index $\ell^{(i)} \in [M]$.

[3]Note that the location information can be easily fed back through lower-frequency links.

### *B. Joint Handover and Beam Tracking Algorithm*

Algorithm 1 describes our proposed joint handover and beam tracking algorithm along a trajectory with duration $M$. If the current association cannot offer the required SNR level, the decision regarding handover or beam track is made based on $a^{(i)}$ as the output of the RL algorithm. In the case of the handover decision, the value of $a^{(i)}$ represents the target BS.

The beam tracking algorithm based on small spatial measurement in time slot $i$ is shown in Algorithm 2. In slot $i$, the algorithm starts by using the main beam of the same serving BS in the previous time slot $i - 1$. If the SNR value is lower than the threshold, then starts a small spatial measurement over the AoD direction of the main beam. To quantify the size of the spatial neighbourhood, we define $\Delta\phi$ and $\Delta\theta$ as the maximum absolute horizontal and vertical deviation from the main AoD direction. We define $\delta\phi$ and $\delta\theta$ as the measurement resolution in horizontal and vertical, respectively. Inspired by [15], the spatial neighbourhood $\mathcal{N}$ surrounding the main AoD direction can be expressed using the horizontal neighbourhood $\mathcal{N}_\phi$ and vertical neighbourhood $\mathcal{N}_\theta$ as

$$\mathcal{N}_\phi(\Delta\phi, \delta\phi) = \left\{i.\delta\phi : i \in \left[-\left\lfloor\frac{\Delta\phi}{\delta\phi}\right\rfloor, \left\lfloor\frac{\Delta\phi}{\delta\phi}\right\rfloor\right]\right\} \quad (8)$$

$$\mathcal{N}_\theta(\Delta\theta, \delta\theta) = \left\{j.\delta\theta : j \in \left[-\left\lfloor\frac{\Delta\theta}{\delta\theta}\right\rfloor, \left\lfloor\frac{\Delta\theta}{\delta\theta}\right\rfloor\right]\right\} \quad (9)$$

where $\lfloor.\rfloor$ is the floor operation. The complete neighbourhood is the Cartesian product of the horizontal and vertical neighbourhoods as

$$\begin{aligned}\mathcal{N}(\Delta\phi, \Delta\theta, \delta\phi, \delta\theta) &= \mathcal{N}_\phi(\Delta\phi, \delta\phi) \times \mathcal{N}_\theta(\Delta\theta, \delta\theta) \\ &= \{(\phi, \theta) : \phi \in \mathcal{N}_\phi(\Delta\phi, \delta\phi), \theta \in \mathcal{N}_\theta(\Delta\theta, \delta\theta)\} \quad (10)\end{aligned}$$

The spatial neighborhoods $\mathcal{T}^{(i)}$ in time slot $i$ surrounding the main AoD directions $(\phi_{\ell^\star}^{(i-1)}, \theta_{\ell^\star}^{(i-1)})$ in previous time slot is

$$\mathcal{T}^{(i)} = (\phi_{\ell^\star}^{(i-1)}, \theta_{\ell^\star}^{(i-1)}, ) + \mathcal{N}(\Delta\phi, \Delta\theta, \delta\phi, \delta\theta). \quad (11)$$

Now given the main AoD direction, we need to find the transmit direction from neighbourhoods $\mathcal{T}^{(i)}$ that offers the SNR threshold. We represent the sorted direction pairs as $[\mathcal{T}^{(i)}]_\mathcal{I}$, where $\mathcal{I}$ is the sorted indices. It means the directions in $[\mathcal{T}^{(i)}]_\mathcal{I}$ increase in distance from the main AoD direction. Starting from the main AoD direction, the SNR of each transmit direction in $[\mathcal{T}^{(i)}]_\mathcal{I}$ is measured until a beam pair meets the required SNR level. Afterwards, no further measurements are required. If no direction meets the threshold, the entire $(\Delta\phi, \Delta\theta)$-neighbourhood is measured to find the beam pairs that offer the SNR threshold.

Note that in the worse scenario, if the selected target BS based on our proposed algorithm cannot offer the required SNR level due to very sudden blockage, the conventional handover methods based on searching over the candidate BSs in UEs vicinity can be applied. However, as shown in the numerical results, such extreme case is rare.

**Algorithm 1** Joint handover and beam tracking

**Input**: Trajectory with duration $M$

Initialization: for $i = 1$ set $j_S^{(1)}$=1
**for** $i \in 1, ..., M$ **do**
  **if** $\text{SNR}_{j_S}^{(i)} < \text{SNR}_{\text{thr}}$ **then**
    Choose the optimal action $a^{(i)}$ based on current $s^{(i)}$.
    **if** $a^{(i)} \neq 0$ **then**. ▷ handover execution
      Set $j_S^{(i)} = a^{(i)}$ and run the initial beam training process and compute the achieved throughput $\Gamma^{(i)}$ as (3).
    **else**
      Run Algorithm 2 and compute $\Gamma^{(i)}$.
    **end if**
  **end if**
**end for**

**Output**: $\Gamma^{(i)}$

**Algorithm 2** Beam tracking in time slot $i$ at the BS $j$

**Input**: $[\mathcal{T}^{(i)}]_{\mathcal{I}}$, $\text{SNR}_{\text{thr}}$, duration of each beam pair testing ($\beta$), $\text{cnt}^{(i)} = 0$.

**for** $(\phi, \theta) \in [\mathcal{T}]_{\mathcal{I}}$ **do**
  Set $\mathbf{f}_j^{(i)} = \mathbf{a}(\phi, \theta)$.
  Measure $\text{SNR}_j^{(i)}$ as (2).
  Set $\text{cnt}^{(i)} = \text{cnt}^{(i)} + 1$. ▷ number of beam pair testing
  **if** $\text{SNR}_j^{(i)} >= \text{SNR}_{\text{thr}}$ **then**
    $(\phi_{\ell^\star}^{(i)}, \theta_{\ell^\star}^{(i)}) = (\phi^{\text{BS}}, \theta^{\text{BS}})$
    $\tau_b^{(i)} = \beta.\text{cnt}^{(i)}$
    **break;**
  **end if**
**end for**

compute the achieved throughput $\Gamma^{(i)}$ as (3)

## IV. Numerical Results

We evaluate the performance of the proposed method in an urban environment using the ray tracing tool in the MATLAB toolbox. The output of the ray tracing tool is the $L$ available paths between a BS and a UE in a specific location. The ray tracing maintains the spatial consistency of mmWave channels. As depicted in Fig. 2, we extracted the building map of Kista in Stockholm city, Sweden and used it as the input data for the ray tracing simulation. In our scenario, we assumed the building material is *brick* and the terrain material is *concrete*. We also add some random obstacles in the street with different heights (1 m and 3 m) and widths (2 m and 4 m) as the human bodies and various vehicles. These temporary obstacles are distributed randomly in the street with density $10^{-2}$ per $m^2$. The material loss and the location of the temporary obstacles are chosen randomly in each realization of the channel. The BSs are located on the wall of buildings. The location of the BSs is chosen randomly while covering the entire trajectory. The BSs' height is 6 m. We consider a pedestrian mobility model with a speed of 1 m/s. We consider the different lengths of the trajectories as $100T_{\text{A}}$, $200T_{\text{A}}$, $300T_{\text{A}}$, $400T_{\text{A}}$, $500T_{\text{A}}$. The main simulation parameters are listed in Table I.

Fig. 2: Simulation area in Kista, Stockholm. The yellow line shows the trajectory. Stars show the location of the BSs.

In the simulation, we consider the $\text{SNR}_{\text{thr}} = 2$ dB and the throughput threshold $\Gamma_{\text{thr}} = 1$ bit/Hz. The value of $\tau_c$ is 10 $ms$. In the case of handover, we fix the initial beam training duration as $\tau_b = \frac{1}{3}\tau_c$. In the case of beam tracking, $\tau_b$ is not fixed and equals the size of measuring neighbourhood multiplied by the duration of each beam pair testing ($\beta = 10$ $\mu s$). We compare the performance of our proposed method with two baselines. To have a fair comparison, we choose two baselines in which the target BS for the handover is pre-determined. Hence, we do not take into account the discovery time of finding the target BS in the baselines. Just like in our method, the handover is triggered if $\text{SNR} < \text{SNR}_{\text{thr}}$.

As **Baseline 1** we consider the multi-connectivity method [8]. We implement a scenario where the UE maintains its connection with a nearby BS as a backup solution while being connected to the serving BS and once it experiences the blockage of the serving link, starts connecting to the backup solution. As **Baseline 2** we select the learning-based handover in [11]. The method shows very good performance in maximizing the achieved rate along the trajectory. In this baseline, the target BS during the handover process is determined by a learning algorithm. Although the target BSs are selected based on the long-term effect on the achieved rate, still can cause frequent handovers and throughput degradation.

First, we fix the number of BSs to 10 (see Fig. 2). We consider $10^4$ different channel realization as the input of the RL algorithm. After getting the optimal policy, we test it over real-time measurements and report the average of the performance over 500 channel realizations. Fig. 3 shows the average number of locations with unmet throughput thresholds along the trajectory with different lengths and Fig. 4 shows the average number of handovers needed. In comparison to the other two baselines, our method provides better throughput results by selecting to perform either beam tracking or a handover. Furthermore, we note that the two baselines have a higher number of handovers than our method due to only considering the handover solution. Hence, by considering the joint handover and beam tracking problem our method provides better-achieved throughput while decreasing the number of handovers. Fig. 5 shows the average aggregate achieved

Table I: Simulation parameters.

| Parameters | Values in Simulations |
|---|---|
| BS transmit power | 10 dBm |
| Noise power level | $\sigma^2$=-174 dBm/Hz |
| Signal bandwidth | 100 MHz |
| BS antenna | $8 \times 8$ uniform planar array [11] |
| Time interval duration | $T_{\mathrm{A}} = 1s$ |
| Neighborhood size | $(\Delta\phi, \Delta\theta) = (10°, 10°)$ |
| Measurement resolution | $(\delta\phi, \delta\theta) = (5°, 5°)$ |
| Discount factor | $\eta = 0.99$ |
| $\lambda$ | 100 |

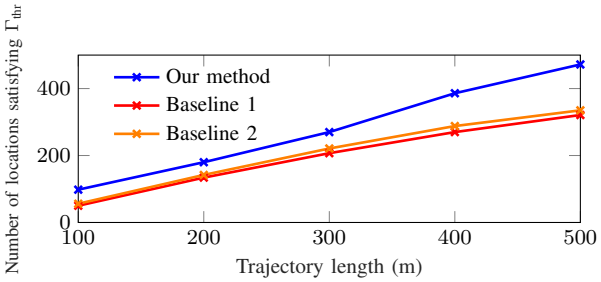


Fig. 3: The average number of locations with unmet throughput threshold for different lengths of the trajectory.

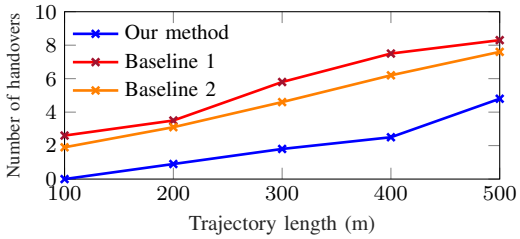


Fig. 4: The average number of handovers for different lengths of the trajectory.

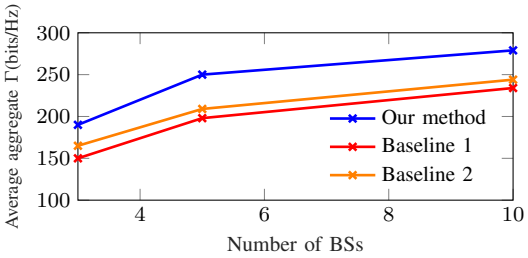


Fig. 5: The average aggregate achieved throughput per Hz along the trajectory with length 300 m.

throughput along the trajectory with length 300 m for different numbers of BSs. By increasing the number of BSs the number of the locations satisfying the $\Gamma_{\mathrm{thr}}$ also increases hence the aggregate throughput along the trajectory increases. Even with a small number of BSs, our method outperforms baselines in aggregate throughput along the trajectory by determining whether to use a handover or beam tracking solution.

We consider 10000 iterations during the training in our method and Baseline 2. With the training machine MacBook Pro 2020 M1 with a memory of 16 GB, each iteration takes about 15 seconds. Note that the absolute value of the training time per iteration depends on the running machine.

## V. Conclusions

In this work, we proposed and studied a learning-based joint handover and beam tracking method in a mobile mmWave network. The aim of our algorithm is to maximize the aggregate throughput of the UE along a trajectory and ensure the achieved throughput in each location is higher than the threshold. Our evaluation results showed that by making an optimal decision regarding handover execution or beam tracking, our method provides high achievable throughput and reduces the number of handovers. Considering different mobility models and studying the effect of neighbouring size can be valuable future work.

## References


[1] T. S. Rappaport, S. Sun, R. Mayzus, H. Zhao, Y. Azar, K. Wang, G. N. Wong, J. K. Schulz, M. Samimi, and F. Gutierrez Jr, "Millimeter wave mobile communications for 5G cellular: It will work!" *IEEE Access*, vol. 1, no. 1, pp. 335–349, May 2013.

[2] H. Hassanieh, O. Abari, M. Rodriguez, M. Abdelghany, D. Katabi, and P. Indyk, "Fast millimeter wave beam alignment," in *Pro. ACM SIGCOM*, 2018, pp. 432–445.

[3] 3GPP, "Requirements for support of radio resource management," *Standard 3GPP TS 38.138*, no. TS 36.133, v15.19.0, Sep. 2022.

[4] R. W. Heath, N. Gonzalez-Prelcic, S. Rangan, W. Roh, and A. M. Sayeed, "An overview of signal processing techniques for millimeter wave mimo systems," *IEEE J. Sel. Top. Signal Process.*, vol. 10, no. 3, pp. 436–453, Apr. 2016.

[5] X. Sun, C. Qi, and G. Y. Li, "Beam training and allocation for multiuser millimeter wave massive mimo systems," *IEEE Trans. Wirel. Commun.*, vol. 18, no. 2, pp. 1041–1053, 2019.

[6] D. Zhang, S. Shen, C. She, M. Xiao, Z. Pang, Y. Li, and L. Wang, "Training beam sequence design for mmwave tracking systems with and without environmental knowledge," *IEEE Trans. Wirel. Commun.*, 2022.

[7] M. F. Özkoç, A. Koutsaftis, R. Kumar, P. Liu, and S. S. Panwar, "The impact of multi-connectivity and handover constraints on millimeter wave and terahertz cellular networks," *IEEE J-SAC.*, vol. 39, no. 6, pp. 1833–1853, 2021.

[8] M. Gapeyenko, V. Petrov, D. Moltchanov, M. R. Akdeniz, S. Andreev, N. Himayat, and Y. Koucheryavy, "On the degree of multi-connectivity in 5G millimeter-wave cellular urban deployments," *IEEE Trans. Veh. Technol.*, vol. 68, no. 2, pp. 1973–1978, Feb. 2019.

[9] "Multi-connectivity; overall description," *Standard 3GPP*, vol. v16.1.0, no. TS 37.340, 2020.

[10] Y. Sun, G. Feng, S. Qin, Y. Liang, and T. P. Yum, "The smart handoff policy for millimeter wave heterogeneous cellular networks," *IEEE Trans Mob Comput.*, vol. 17, no. 6, pp. 1456–1468, Jun. 2018.

[11] S. Khosravi, H. S. Ghadikolaei, and M. Petrova, "Learning-based handover in mobile millimeter-wave networks," *IEEE TCCN*, vol. 7, no. 2, pp. 663–674, 2021.

[12] A. Patra, L. Simić, and P. Mähönen, "Smart mm-wave beam steering algorithm for fast link re-establishment under node mobility in 60 ghz indoor wlans," in *Proceedings of the 13th ACM International Symposium on Mobility Management and Wireless Access*, 2015, pp. 53–62.

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

[14] D. Bertsekas, *Reinforcement Learning and optimal control*. Athena Scientific, 2019.

[15] I. P. Roberts, A. Chopra, T. Novlan, S. Vishwanath, and J. G. Andrews, "Steer: Beam selection for full-duplex millimeter wave communication systems," *IEEE Trans Commun.*, pp. 1–1, 2022.