

An Online Context-aware Machine Learning Algorithm for 5G mmWave Vehicular Communications

Gek Hong (Allyson) Sim*, Sabrina Klos (née Müller)[†], Arash Asadi*, Anja Klein[†], Matthias Hollick*

*Secure Mobile Networking lab (SEEMOO), Technische Universität Darmstadt
{firstname.lastname}@seemoo.tu-darmstadt.de

[†]Communication Engineering Lab, Technische Universität Darmstadt
{firstname.lastname}@nt.tu-darmstadt.de

Abstract—Millimeter-Wave (mmWave) bands have become the de-facto candidate for 5G vehicle-to-everything (V2X) since future vehicular systems demand Gbps links to acquire the necessary sensory information for (semi)-autonomous driving. Nevertheless, the directionality of mmWave communications and its susceptibility to blockage raise severe questions on the feasibility of mmWave vehicular communications. The dynamic nature of 5G vehicular scenarios, and the complexity of directional mmWave communication calls for higher context-awareness and adaptability. To this aim, we propose an *online learning algorithm addressing the problem of beam selection with environment-awareness in mmWave vehicular systems*. In particular, we model this problem as a contextual multi-armed bandit problem. Next, we propose a lightweight context-aware online learning algorithm, namely fast machine learning (FML), with proven performance bound and guaranteed convergence. FML exploits coarse user location information and aggregates received data to learn from and adapt to its environment. Furthermore, we demonstrate the feasibility of a real-world implementation of FML by proposing a standard-compliant protocol based on the existing architecture of cellular networks and the forthcoming features of 5G. We also perform an extensive evaluation using realistic traffic patterns derived from Google Maps. Our evaluation shows that FML enables mmWave base stations to achieve near-optimal performance on average within 33 mins of deployment by learning from the available context. Moreover, FML remains within $\sim 5\%$ of the optimal performance by swift adaptation to system changes (i.e., blockage, traffic).

I. INTRODUCTION

Recent studies highlight the necessity of *multi-Gbps links* to enable 5G vehicle-to-everything (V2X) communications [2], [3]. Such a high data rate link is needed to acquire accurate sensory data (e.g., HD maps, radar feeds), which is crucial for (semi)-autonomous driving. Due to high congestion in sub-6GHz bands used by 4G LTE-A systems, the 5G community plans to exploit the underutilized mmWave bands (10-300 GHz). This underutilization is due to the impairments of mmWave bands, such as high pathloss and penetration loss. Nevertheless, new research demonstrates that: (i) directional transmission and beamforming is the solution to compensate

This manuscript is an extended version of the work accepted for presentation at IEEE INFOCOM 2018 [1].

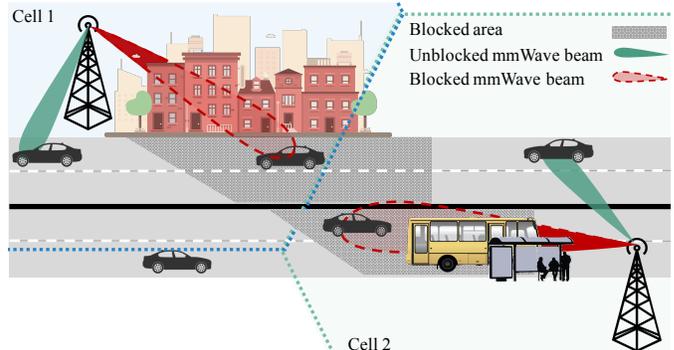


Fig. 1. An example of a mmWave cellular scenario and the impact of different sources of blockage.

for the high pathloss, and (ii) higher deployment density of base stations is the remedy for short communication range in mmWave bands (100-150 m) [4], [5].

These solutions prove the feasibility of mmWave communication. However, they bring about many new challenges in the system design. Firstly, the directional communication requires *accurate beam alignments* between the base station and the vehicle [6], which is unnecessary for the omnidirectional transmission in sub-6GHz bands. Secondly, mmWave signals are *prone to blockages* (e.g., buildings, foliage) due to high penetration loss (see Fig. 1). Thus, the performance of mmWave systems can be severely hampered by inaccurate *beam selection*. The performance degradation can be mitigated by enabling the base stations to perform beam selection based on their surrounding environment (e.g., to avoid blockages). In today's network, this knowledge is populated via on-site signal measurements (e.g., war-driving tests), which are time-consuming and unscalable for dense 5G deployments. Moreover, this approach cannot account for dynamic traffic patterns and blockages. We believe that the base stations should autonomously *explore, learn from, and adapt to* their environment to make accurate beam selection while maintaining sustainable scalability. To date, there is no proposal fostering such a capability at mmWave base stations [7]. To

this aim, a practical approach should allow the base station to characterize its surroundings autonomously by exploiting the available contextual information. In particular, the correlation between this information (e.g., location of the users) and the outcome of a decision (e.g., beam selection) is the key to optimal future decisions. This emphasizes the necessity of autonomous learning more than ever, specifically to cope with the massive densification of 5G networks [8], [9].

FML algorithm. In this paper, we propose fast machine learning (FML), which is a low-complexity and a scalable online learning algorithm for mmWave base stations. FML is coupled with a practical protocol, which is designed based on the features of the forthcoming 5G cellular network. We model the beam selection as a *contextual multi-armed bandit problem* and propose a *contextual online learning algorithm*. This algorithm enables the mmWave base stations to autonomously learn from prior decisions and their relations to the available contextual information. In particular, FML explores different beams over time while accounting for contextual information (i.e., vehicles' direction of arrival). The outcome of the exploration is used to adapt to system dynamics such as the appearance of blockages and changes in traffic patterns. FML identifies blockages by evaluating the aggregate received data of each vehicle for each selected beam. FML also adapts to traffic patterns by learning the correlation between the direction of arrival and the received data. Let's take the mmBS on the right-hand-side in Fig. 2 as an example. On the one hand, when vehicles are coming from the north (i.e., the yellow car), the mmBS will use the beams pointing to the north (as shown by the blue mmWave beams) as it results in a higher throughput. On the other hand, when the vehicles are arriving from the east (i.e., the white car), then the mmBS will select the beams that are pointing to the south (as shown by the orange color beams). As a result, an mmBS that employs FML selects the beams which maximize the aggregate network capacity by accounting for the traffic pattern. Consequently, FML provides higher coverage to the roads with higher traffic and hence, it serves a larger number of vehicles compared to non-contextual algorithms.

FML fights the issues of mmWave vehicular communication in several fronts: (i) it detects permanent blockages (e.g., buildings), and frequently blocked areas due to temporary blockages (e.g., parking spots, bus stations or construction sites frequented by large trucks) using online learning; (ii) it leverages traffic patterns to maximize the system capacity by providing larger coverage (i.e., allocation of more beams) in areas with heavier traffic. This is important because mmWave base stations can transmit simultaneously over a limited number of beams. This limitation depends on the hardware characteristics, the mmWave channel sparsity, and the beamforming technique¹; (iii) it infers traffic patterns from the context (i.e.,

¹To allow simultaneous transmission over all beams with analogue/hybrid beamforming techniques, the number of RF chains should be proportional to the number of beams. Such a design is undesirable both in terms of form-factor and manufacturing cost. Further technical details on this limitation can be found in [5] and [10].

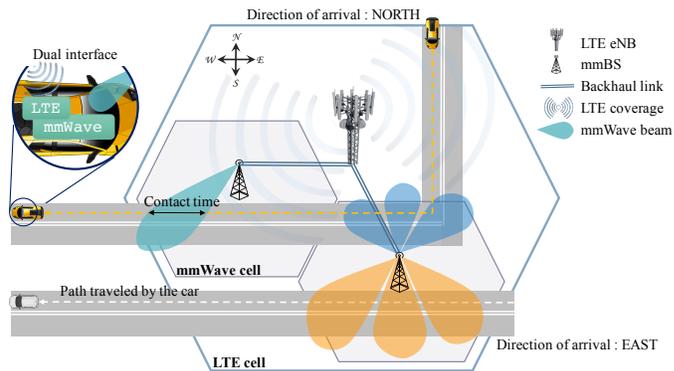


Fig. 2. Illustration of our system model. For clarity, the figure contains only two mmBSs. Each mmBS can transmit over two beams simultaneously. The direction of arrival of the vehicles (shown in dashed line) is derived from the location of the vehicle upon registration to the eNB.

the vehicle's direction of arrival) and selects the best beams. Majority of roads have distinct traffic patterns influenced by the time of the day. For example, the traffic in the main streets moves towards the financial center early in the morning and away from it in the evening (i.e., towards the residential areas). While interpreting these patterns is out of the scope of this paper, we design FML to identify and learn from such patterns. Specifically, FML identifies the change of the traffic patterns through the context information and its impact on the achievable throughput. This change triggers FML to re-explore the performance of the beams and to adapt its beam selection accordingly.

Our contributions. The following summarizes the contributions of this paper:

- We model the beam selection at mmWave base stations as a contextual multi-armed bandit problem. Our model is generic, and it can be easily adapted to different contexts for new 5G use-cases.
- We provide the first contextual online learning algorithm for beam selection in mmWave base stations. The algorithm enables the base stations to autonomously learn the data rate of every beam, without requiring a training phase.
- We give an analytical upper bound on the regret, i.e., the loss of learning, which proves convergence of FML to the optimal beam selection.
- We go beyond theory by illustrating how FML can be practically implemented in a 5G cellular system. In particular, we elaborate on design aspects of the FML and its corresponding signaling requirements from an architectural point of view.
- We demonstrate by means of extensive simulation that – with live and typical traffic patterns obtained from Google Maps at our premises – FML substantially outperforms the benchmark algorithms.

II. SYSTEM MODEL

We consider a heterogeneous cellular system in which mmWave base stations (mmBSs) overlay the coverage area

of an LTE eNB (see Fig. 2). This network model is widely expected for forthcoming 5G systems [11]–[13]. The mmBSs are connected to the eNB via a backhaul link. The vehicles are equipped with (i) an LTE interface to maintain a connection to the eNB, and (ii) an mmWave interface for high-speed data communication. We focus on the downlink in this paper. Note that FML aims to find the best beam for communication, regardless of its direction (uplink or downlink). Hence, our analysis can be applied in uplink since the beam alignment/selection is the same for uplink and downlink. We assume neither the eNB nor the mmBSs have any knowledge of their surroundings. We focus on a system with small signaling overhead. On the one hand, the only information available to the mmBS is the direction of arrival of the vehicle (i.e., north, south, east, west), which we define below formally as the vehicle context. On the other hand, the vehicles will only know the location of the mmBS and the selected beam(s). In Section IV, we elaborate how this information is communicated within the network.

A. Choice of Learning Method

We model beam selection in an mmBS as an online learning problem. This is because it allows the mmBS to identify the best beams autonomously over time while accounting for dynamic traffic and environment changes. Specifically, we model the problem as a *multi-armed bandit problem* (MAB problem). Various problems in wireless communications have been treated using MABs [14]. In MAB problems, a decision maker has to select a subset of actions of unknown expected rewards with the goal to maximize the reward over time [15]. The challenge in MAB problems lies in solving the exploration vs. exploitation dilemma, since all actions should be explored sufficiently often to learn their rewards, but also those actions which have already yielded high rewards should be exploited. We model our problem as a MAB problem since an mmBS may only use a limited set of beams simultaneously (as shown in Fig. 2). Therefore, the mmBS needs to identify the best beams by carefully selecting subsets of beams over time. More specifically, our approach falls under the category of *contextual multi-armed bandit problems*. In contextual MAB problems, the decision maker is first presented with some context information, before selecting an action. This context information affects the rewards of the actions [16]–[19]. The additional challenge in contextual MAB problems is how to exploit historical reward observations under similar contexts. We model our problem as a contextual MAB problem since, in this way, the mmBS does not simply learn which beams are the best *on average*, but instead it exploits additional information about approaching vehicles to identify which beams are the best under a given traffic situation. Then, we propose a contextual online learning algorithm for our problem, which tackles the above-mentioned challenges.

B. Problem Formulation

The mmBS can use a finite set \mathcal{B} of $B = |\mathcal{B}|$ distinct, orthogonal beams (see Fig. 2). We assume that the mmBS

may only select a subset of m beams simultaneously, where $m \in \mathbb{N}$, $m < B$, is a fixed number. This limitation is imposed by the mmWave channel sparsity, beamforming technique, and the hardware characteristics (e.g., number of RF chains) [5]. The goal of the mmBS is to select a subset of m beams that maximizes the amount of data successfully received by the bypassing vehicles in the coverage area. We assume that the mmBS is unaware of its surrounding, i.e., the mmBS does not have prior knowledge about its environment (e.g., street course, blockages). This significantly reduces the complexity of the network implementation as the operator does not need to configure each mmBS based on its surroundings. Hence, the mmBS should learn over time the best subset of beams for its environment. For this purpose, the mmBS should take into account vehicles' context, since the best beams depends on the context of bypassing vehicles (e.g., their directions of arrival).

We consider a discrete time setting, where the mmBS updates its beam selection in regular time periods. In each period $t = 1, \dots, T$, where $T \in \mathbb{N}$ is a finite time horizon, the following events happen:

- (i) A set $\mathcal{V}_t = \{v_{t,i}\}_{i=1,\dots,V_t}$ of $V_t = |\mathcal{V}_t|$ vehicles registers to the mmBS via the LTE eNB. The number of vehicles satisfies $V_t \leq V_{\max}$, where $V_{\max} \in \mathbb{N}$ is the maximum number of supported vehicles in the system, which corresponds to the maximum number of vehicles that fit on the streets within the coverage area of the mmBS. The registration process is described in Section IV. During the registration process, the mmBS receives the required information about the *context* $x_{t,i}$ of each approaching vehicle $v_{t,i}$. Formally, the context $x_{t,i}$ is an X -dimensional vector taken from the bounded context space $\mathcal{X} = [0, 1]^X$, where we assume that the information about a vehicle is described using X context dimensions. In each of the X dimensions, the context information is encoded as a value between 0 and 1. This generic model allows to model both continuous as well as discrete types of context information. In the numerical simulations of this manuscript, the context vector is one-dimensional (i.e., $X = 1$) since we only consider the directional of arrival as the context. This will be further clarified in Sections IV and V.
- (ii) The mmBS selects a subset of m beams. We denote the set of selected beams in period t by $\mathcal{S}_t = \{s_{t,j}\}_{j=1,\dots,m} \subseteq \mathcal{B}$. Then, the vehicles in \mathcal{V}_t are informed about the selected beams through the associated eNB via their LTE interface.
- (iii) When vehicle $v_{t,i}$ reaches the mmBS's coverage area, the mmBS transmits data to vehicle $v_{t,i}$ and observes the amount of data $r_{s_{t,j}}(x_{t,i}, t)$ vehicle $v_{t,i}$ successfully receives via the selected beams $s_{t,j}$, $j = 1, \dots, m$, until the end of the period t .

In general, the amount of data $r_b(x)$ a vehicle with context $x \in \mathcal{X}$ can successfully receive from the mmBS using beam $b \in \mathcal{B}$ during one period is a random variable that depends

on the environment of the mmBS (e.g., street conditions and course, blockages, etc.). We call the random variable $r_b(x)$ the *beam performance* (i.e., the aggregate received data by the vehicle) of beam b under context x . We assume that this random variable is bounded in $[0, R_{\max}]$, where R_{\max} is the maximum amount of data that can be received by a vehicle. R_{\max} is bounded by the maximum achievable rate of the channel, and it depends on the selected modulation and coding scheme. The contact time as shown in Fig. 2 (i.e., the time within which mmBS can transmit data to the vehicle) is bounded by the coverage area of the beam, which depends on the beam width and speed. By $\mu_b(x)$, we denote the expected value of random variable $r_b(x)$, and we call it the *expected beam performance* of beam b under context x . The mmBS aims at selecting a subset of beams which maximizes the expected received data at the vehicles, i.e., maximizes the sum of the expected beam performances. We denote the optimal subset in period t , by $\mathcal{B}_t^*(\mathcal{X}_t) = \{b_{t,j}^*(\mathcal{X}_t)\}_{j=1,\dots,m} \subseteq \mathcal{B}$. The set $\mathcal{B}_t^*(\mathcal{X}_t)$ depends on $\mathcal{X}_t = \{x_{t,i}\}_{i=1,\dots,V_t}$ and its m beams formally satisfy

$$b_{t,j}^*(\mathcal{X}_t) \in \underset{b \in \mathcal{B} \setminus (\cup_{k=1}^{j-1} \{b_{t,k}^*(\mathcal{X}_t)\})}{\operatorname{argmax}} \sum_{i=1}^{V_t} \mu_b(x_{t,i}) \quad (1)$$

for $j = 1, \dots, m$. Hence, if the mmBS *knew* the expected beam performances $\mu_b(x)$ for each vehicle context $x \in \mathcal{X}$ and each beam $b \in \mathcal{B}$ *a priori, like an oracle*, it could simply select the optimal subset of beams for each set of approaching vehicles according to (1). Over the sequence $1, \dots, T$ of periods, this would yield an expected amount of

$$\sum_{t=1}^T \sum_{i=1}^{V_t} \sum_{j=1}^m \mathbb{E}[r_{b_{t,j}^*(\mathcal{X}_t)}(x_{t,i})] = \sum_{t=1}^T \sum_{i=1}^{V_t} \sum_{j=1}^m \mu_{b_{t,j}^*(\mathcal{X}_t)}(x_{t,i}) \quad (2)$$

data that can be received in total.

However, the mmBS does not know the environment, and hence it has to learn the expected beam performances $\mu_b(x)$ over time. In order to learn these values, the mmBS has to try out different beams for different vehicle contexts over time. At the same time, it should ensure that those beams that were already proven to be good are used sufficiently often. Hence, the mmBS has to find a trade-off between exploring beams of which it has little knowledge and exploiting beams with high average beam performance. In the next section, we will present a *learning algorithm*, which for each period with approaching vehicles of contexts \mathcal{X}_t , selects a subset \mathcal{S}_t of m beams. The selection of the learning algorithm depends on the history of selected beams in previous periods and the corresponding observed beam performances. Given an arbitrary sequence of vehicle arrivals with arbitrary contexts the expected amount of data received by the vehicles is given by:

$$\sum_{t=1}^T \sum_{i=1}^{V_t} \sum_{j=1}^m \mathbb{E}[r_{s_{t,j}}(\mathcal{X}_t)(x_{t,i})] = \sum_{t=1}^T \sum_{i=1}^{V_t} \sum_{j=1}^m \mathbb{E}[\mu_{s_{t,j}}(\mathcal{X}_t)(x_{t,i})]. \quad (3)$$

In (3), the expectation is taken with respect to the selections of the learning algorithm and the randomness of beam performances.

The expected difference in the amount of received data achieved by an oracle and by the learning algorithm is called the *regret of learning*. Given (2) and (3), it is defined as:

$$R(T) = \mathbb{E} \left[\sum_{t=1}^T \sum_{i=1}^{V_t} \sum_{j=1}^m \left(r_{b_{t,j}^*(\mathcal{X}_t)}(x_{t,i}) - r_{s_{t,j}}(\mathcal{X}_t)(x_{t,i}) \right) \right] \\ = \sum_{t=1}^T \sum_{i=1}^{V_t} \sum_{j=1}^m \left(\mu_{b_{t,j}^*(\mathcal{X}_t)}(x_{t,i}) - \mathbb{E}[\mu_{s_{t,j}}(\mathcal{X}_t)(x_{t,i})] \right). \quad (4)$$

III. FML ALGORITHM

The above problem formulation corresponds to a contextual multi-armed bandit problem and we propose a contextual online learning algorithm inspired by [19]. Intuitively, the algorithm learns the expected beam performances under different contexts online over time. The algorithm works on the assumption that for similar vehicle contexts, the performance of a particular beam will on average be similar.

The algorithm first uniformly partitions the context space into small sets of similar contexts and learns about the performance of different beams independently in each of these small sets. Then, in each of its discrete periods, the algorithm either enters an exploration phase or an exploitation phase. The phase it enters is decided based on the contexts of approaching vehicles and based on a control function. While in exploration phases, the algorithm selects a random subset of beams, in exploitation phases, the algorithm selects beams that showed the highest performance when selected in previous periods. By observing the amount of data received by vehicles in the system, the algorithm acquires performance estimates of beams; thereby, it learns the performance of the different beams under different vehicle contexts over time.

A. Detailed Description

In detail, our proposed beam selection algorithm, called FML works as follows (see Algorithm. 1): First, during initialization (lines 2-4), FML uniformly partitions the context space $\mathcal{X} = [0, 1]^X$ into $(p_T)^X$ X -dimensional hypercubes of size $(\frac{1}{p_T})^X$, where p_T is an input to the algorithm. We call the resulting partition \mathcal{P}_T . Moreover, FML initializes the counters $N_{b,h}(t)$ for each beam $b \in \mathcal{B}$ and each hypercube $h \in \mathcal{P}_T$. Intuitively, these counters are used to describe how many vehicles of a certain context have already arrived at the mmBS in previous periods, in which the mmBS had selected a certain beam. Formally, the counter $N_{b,h}(t)$ represents the total number of vehicles with the context in hypercube h that approached the mmBS whenever beam b had been selected in any of the periods $1, \dots, t-1$. In addition, the algorithm initializes the estimates $\hat{\mu}_{b,h}(t)$ for each beam $b \in \mathcal{B}$ and each hypercube $h \in \mathcal{P}_T$. The estimator $\hat{\mu}_{b,h}(t)$ represents the estimated beam performance of beam b for vehicles with the context in hypercube h .

In period t , FML observes the contexts $\mathcal{X}_t := \{x_{t,i}\}_{i=1,\dots,V_t}$ of the V_t approaching vehicles and for each context $x_{t,i}$, FML determines to which hypercube this context belongs to (lines 6-7), i.e., it finds $h_{t,i} \in \mathcal{P}_T$ with $x_{t,i} \in h_{t,i}$. Based on the collection $\mathcal{H}_t := \{h_{t,i}\}_{i=1,\dots,V_t}$ of hypercubes, FML next (line 8) computes the set $\mathcal{B}_{\mathcal{H}_t}^{\text{ue}}(t)$ of under-explored beams via

$$\mathcal{B}_{\mathcal{H}_t}^{\text{ue}}(t) := \cup_{i=1}^{V_t} \{b \in \mathcal{B} : N_{b,h_{t,i}}(t) \leq K(t)\}, \quad (5)$$

where $K : \{1, \dots, T\} \mapsto \mathbb{R}$ is a deterministic, monotonically increasing control function, which the algorithm gets as input. The control function $K(t)$ is used to decide whether to enter an exploration or an exploitation phase. The control function needs to be chosen adequately in order to guarantee that FML achieves a good performance in terms of its regret. In Theorem 1, we provide a suitable choice for the control function. If there are under-explored beams, FML enters an exploration phase (lines 9-16). In case the number $u(t) := |\mathcal{B}_{\mathcal{H}_t}^{\text{ue}}(t)|$ of under-explored beams is at least m , FML randomly selects m of them. In case the number $u(t)$ of under-explored beams is smaller than m , FML selects all $u(t)$ beams. In addition, it selects the $(m - u(t))$ beams $\hat{b}_{1,\mathcal{H}_t}(t), \dots, \hat{b}_{m-u,\mathcal{H}_t}(t)$ from $\mathcal{B} \setminus \mathcal{B}_{\mathcal{H}_t}^{\text{ue}}(t)$, which satisfy

$$\hat{b}_{j,\mathcal{H}_t}(t) \in \underset{b \in \mathcal{B} \setminus (\mathcal{B}_{\mathcal{H}_t}^{\text{ue}}(t) \cup \bigcup_{k=1}^{j-1} \{\hat{b}_{k,\mathcal{H}_t}(t)\})}{\operatorname{argmax}} \sum_{i=1}^{V_t} \hat{\mu}_{b,h_{t,i}}(t) \quad (6)$$

for $j = 1, \dots, m - u(t)$. If there are no under-explored beams, FML enters an exploitation phase (lines 17-19). It selects the m beams $\hat{b}_{1,\mathcal{H}_t}(t), \dots, \hat{b}_{m,\mathcal{H}_t}(t)$ from \mathcal{B} , which satisfy

$$\hat{b}_{j,\mathcal{H}_t}(t) \in \underset{b \in \mathcal{B} \setminus (\bigcup_{k=1}^{j-1} \{\hat{b}_{k,\mathcal{H}_t}(t)\})}{\operatorname{argmax}} \sum_{i=1}^{V_t} \hat{\mu}_{b,h_{t,i}}(t) \quad (7)$$

for $j = 1, \dots, m$. After beam selection, FML observes the beam performance of each selected beam for each vehicle within this period (line 20). Using these observations, FML updates its internal counters (lines 21-25).

B. Regret and Choice of Parameters

The regret of FML in (4) can be bounded from above. The upper bound given below is based on the following assumption, which states that, the expected beam performance of a particular beam is similar in similar contexts:

Assumption 1. *There exist $L > 0$ and $\alpha > 0$ such that for all $b \in \mathcal{B}$ and for all $x, y \in \mathcal{X}$, it holds that*

$$|\mu_b(x) - \mu_b(y)| \leq L \|x - y\|^\alpha,$$

where $\|\cdot\|$ denotes the Euclidean norm in \mathbb{R}^X .

The regret of FML can be bounded as follows (see [19])

Theorem 1 (Bound for $R(T)$). *Let $K(t) = t^{\frac{2\alpha}{3\alpha+X}} \log(t)$ and $p_T = \lceil T^{\frac{1}{3\alpha+X}} \rceil$. If FML is executed using these parameters and if Assumption 1 holds true, the leading order of the regret $R(T)$ is $O\left(m V_{\max} R_{\max} B T^{\frac{2\alpha+X}{3\alpha+X}} \log(T)\right)$.*

Algorithm 1 Pseudocode of FML algorithm.

```

1: Input:  $T, p_T, K(t)$ 
2: Initialize context partition: Create partition  $\mathcal{P}_T$  of context space  $[0, 1]^X$ 
   into  $(p_T)^X$  hypercubes of identical size
3: Initialize counters: For all  $b \in \mathcal{B}$  and all  $h \in \mathcal{P}_T$ , set  $N_{b,h} = 0$ 
4: Initialize estimates: For all  $b \in \mathcal{B}$  and all  $h \in \mathcal{P}_T$ , set  $\hat{\mu}_{b,h} = 0$ 
5: for each  $t = 1, \dots, T$  do
6:   Observe vehicle contexts  $\mathcal{X}_t = \{x_{t,i}\}_{i=1,\dots,V_t}$ 
7:   Find  $\mathcal{H}_t = \{h_{t,i}\}_{i=1,\dots,V_t}$  such that  $x_{t,i} \in h_{t,i} \in \mathcal{P}_T, i = 1, \dots, V_t$ 
8:   Compute the set of under-explored beams  $\mathcal{B}_{\mathcal{H}_t}^{\text{ue}}(t)$  in (5)
9:   if  $\mathcal{B}_{\mathcal{H}_t}^{\text{ue}}(t) \neq \emptyset$  then ▷ Exploration
10:      $u = \text{size}(\mathcal{B}_{\mathcal{H}_t}^{\text{ue}}(t))$ 
11:     if  $u \geq m$  then
12:       Select  $s_{t,1}, \dots, s_{t,m}$  randomly from  $\mathcal{B}_{\mathcal{H}_t}^{\text{ue}}(t)$ 
13:     else
14:       Select  $s_{t,1}, \dots, s_{t,u}$  as the  $u$  beams from  $\mathcal{B}_{\mathcal{H}_t}^{\text{ue}}(t)$ 
15:       Select  $s_{t,u+1}, \dots, s_{t,m}$  as the  $(m - u)$  beams
        $\hat{b}_{1,\mathcal{H}_t}(t), \dots, \hat{b}_{m-u,\mathcal{H}_t}(t)$  from (6)
16:     end if
17:     ▷ Exploitation
18:     Select  $s_{t,1}, \dots, s_{t,m}$  as the  $m$  beams  $\hat{b}_{1,\mathcal{H}_t}(t), \dots, \hat{b}_{m,\mathcal{H}_t}(t)$ 
       from (7)
19:     end if
20:     Observe received data  $r_{j,i}$  of each vehicle  $v_{t,i}, i = 1, \dots, V_t$ , in
       each beam  $s_{t,j}, j = 1, \dots, m$ 
21:     for  $i = 1, \dots, V_t$  do
22:       for  $j = 1, \dots, m$  do
23:          $\hat{\mu}_{s_{t,j},h_{t,i}} = \frac{\hat{\mu}_{s_{t,j},h_{t,i}} N_{s_{t,j},h_{t,i}} + r_{j,i}}{N_{s_{t,j},h_{t,i}} + 1}$  and  $N_{s_{t,j},h_{t,i}} =$ 
            $N_{s_{t,j},h_{t,i}} + 1$ 
24:       end for
25:     end for
26: end for

```

Proof: The detailed proof of Theorem 1 can be found in [19]. For brevity, we will not repeat the proof here. Essentially, the regret is bounded as follows. First, the regret is divided into three summands. One summand corresponds to the regret due to exploration phases. The other two summands correspond to the regret due to sub-optimal and near-optimal choices in exploitation phases, respectively. Then, each of the three summands is bounded separately. ■

This theorem shows that the regret of FML is sublinear in the time horizon T , i.e., $R(T) = O(T^\gamma)$ with $\gamma < 1$. This implies that $\lim_{T \rightarrow \infty} \frac{R(T)}{T} = 0$ holds, which guarantees that the algorithm has an asymptotically optimal performance. Hence, over time, FML converges to the optimal beam selection strategy. Moreover, for finite time horizon T , the upper bound on the regret characterizes FML's speed of convergence.

IV. IMPLEMENTATION FEASIBILITY OF FML

This section sheds light on the feasibility and practicality of the implementation of FML within the existing 4G cellular architecture and the forthcoming features of 5G systems. In particular, we describe control and data communication within the network. We also elaborate on the procedure that enables FML for 5G V2X communications. In our proposed system, vehicles maintain continuous connectivity via their LTE interface. A vehicle sends a request to attach to an mmBS when mmWave connectivity is needed (e.g., new HD map, local traffic data). In the following, we describe FML in four stages as shown in Fig. 3.

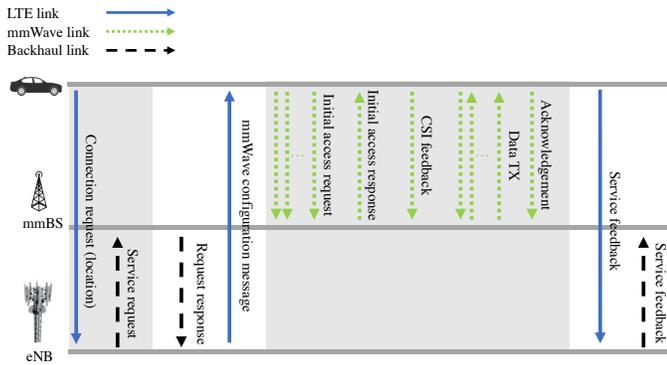


Fig. 3. Integration feasibility of FML in a cellular system.

A. Registration request.

The vehicle sends an mmWave registration request to the serving eNB which contains the vehicle's location. This triggers the eNB to send a service request message to a potential mmBS concerning the vehicle's location. This message contains the vehicle's cellular identifier (e.g., RNTI), and its expected direction of arrival at the mmBS. The conversion of the GPS location to the low-resolution direction of arrival reduces the backhaul signaling overhead. On the one hand, the information exchange for the direction of arrival occurs only once per vehicle. On the other hand, exchange of GPS location requires a continuous coordinate update, which also results in the increase of context space. As we will see in Section V, the performance reduction due to the localization accuracy is negligible as FML always remains very close to the optimal solution.

B. mmBS association.

The mmBS responds to the service request with the information regarding the selected beams. Next, the eNB forwards the mmBS related information (i.e., the location of the mmBS and selected beams) to the vehicle. FML does not require the vehicle to react to this information. However, we argue that knowing the context, the vehicle can perform a simple geometric operation onboard to estimate its arrival to the coverage area of a beam selected by the mmBS. This is advantageous for both omni-directional and directional reception. Firstly, the vehicle only starts listening to the mmWave channel shortly before arrival to the coverage area (lower energy consumption). Secondly, beam-alignment is simplified since the vehicle knows the exact location of the transmission source. Without this knowledge, the beam-alignment for moving objects can be very challenging. As a result, the association becomes more energy efficient and requires less signaling overhead. Considering that the connection setup with a complete 360°-sweep for 802.11ad take a few milliseconds, limiting the scanning angle should reduce this delay to a negligible value.

C. Communication.

Once in coverage, the vehicle starts the regular cellular attachment process by sending an initial access request, which is replied to by a response from the mmBS. The vehicle

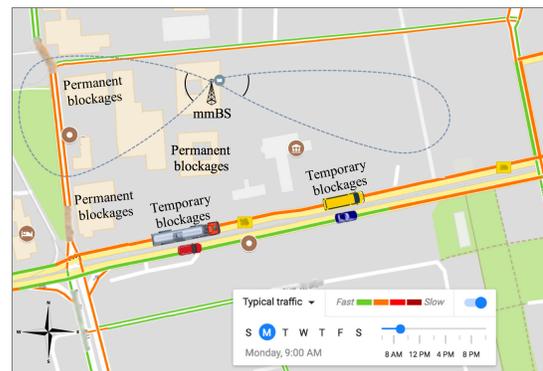


Fig. 4. The map of the simulation environment. The tram and bus temporarily block the red vehicle and blue vehicle, respectively.

measures the channel state information (CSI) from the initial access response message and sends the CSI feedback for modulation and coding assignment. Next, the mmBS starts the data transmission process. Note that although an mmBS may be able to transmit from several beams simultaneously, each beam transmits a separate data-stream (i.e., unicast).

D. Feedback.

No feedback is required if the communication phase is successful because the mmBS has already received acknowledgments for the transmitted frames. In case a vehicle fails to detect the mmBS within a selected beam, it sends the feedback to the serving eNB. This feedback will be forwarded to the mmBS to update FML's future decisions.

V. NUMERICAL EVALUATION

Here, we evaluate and benchmark FML via numerical simulations. In the following, we first describe the simulation environment and the relevant parameters. Next, we provide the pathloss model and other simulation settings, which are chosen according to the 3GPP technical specification in [20]. Then, the benchmark algorithms and results are presented.

A. Simulation Setup

The simulation scenario (e.g., blockages, roads, and traffic patterns) is designed with reference to information obtained from Google Maps in the vicinity of our premises. The mmBS is assumed to have 16 orthogonal beams with variable beam width from 10° to 40° covering the 360° azimuth. The beams are selected according to the recent measurements in [21]. The vehicles enter the system with an arrival rate of λ (in vehicles per second) and their speed varies between 20km/h and 70km/h. Each vehicle chooses one of the routes on the map, whose probability is determined by the typical traffic observed within the area in Google Maps (see Fig. 4). We consider two types of blockages: permanent and temporary blockages. Permanent blockages are the buildings that *permanently* block the path between the mmBS and the vehicles on the road. The temporary blockages (e.g., other vehicles or trams) are modeled by random appearance of objects on either side of the road causing a *temporary* signal blockage. Note that both types of blockages are present in all our evaluations.

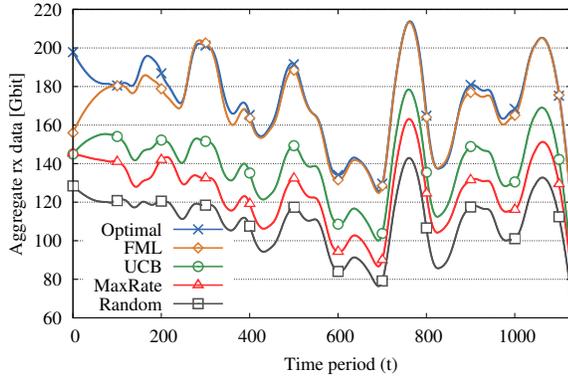


Fig. 5. Aggregate received data for arrival rate with $\lambda = 0.4$ and $m = 4$.

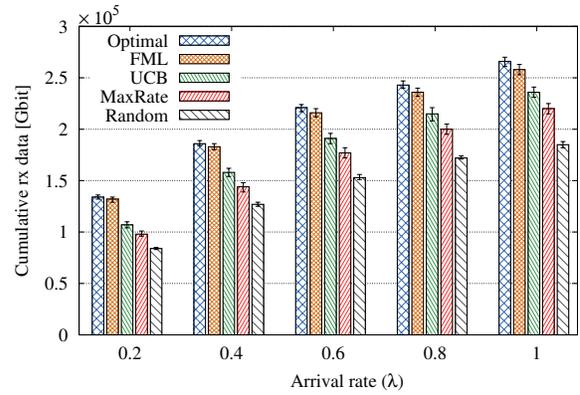


Fig. 6. Impact of arrival rate λ on cumulative received data for $m = 4$ after.

In our implementation, a time period is defined as the time in which the vehicle under observation enters and leaves the cell coverage area of an mmBS. Within this time period, the learning algorithms additionally learn from the context and received data of the other vehicles passing through the selected beams. In this way, we ensure that the algorithms have enough samples to learn from. We choose the direction of arrival as context (i.e., north, south, east, and west), hence the context vector is a one dimensional vector (i.e., $X = 1$). Table I contains the important simulation parameters.

TABLE I
CHANNEL PARAMETERS AS SPECIFIED IN [20].

Parameter	Value
Carrier frequency	28GHz
System Bandwidth	1GHz
Transmit power	30dBm
Noise figure	4dB@mmBS, 7dB@Vehicle
Vehicle's beam width	30°
Thermal noise	-174dBm/Hz
Pathloss model (dB)	$32.4 + 17.3 \log_{10}(d(m)) + 20 \log_{10}(f_c(\text{GHz})) + \xi$ $\xi \sim \mathcal{N}(0, \sigma), \sigma = 1.1\text{dB}$

B. Benchmark Algorithms and Metrics

We provide a thorough performance analysis by comparing FML to several other schemes. The following elaborates on each benchmark scheme:

- **Optimal.** This algorithm has a priori knowledge about the expected beam performance $\mu_b(x)$ of each beam $b \in \mathcal{B}$ in each context $x \in \mathcal{X}$. In each period, Optimal selects the optimal subset $\mathcal{B}_t^*(\mathcal{X}_t)$ of m beams as in (1). Therefore the results achieved by Optimal is the expected performance upper bound of the system.
- **UCB.** This is a variant of the classical learning algorithm UCB [15], which we adapted to our use-case. It learns from previously observed beam performance, but without taking into account the context information obtained from the environment. In each period, UCB selects m beams with the highest estimated upper confidence bounds on their expected beam performance.
- **MaxRate.** This algorithm first explores each beam once. Next, it will select the beam that achieved the highest received data, i.e., the best beam. Once MaxRate selects the best beam, it retains this selection even if the traffic or environmental changes at a later time. MaxRate is

intended to show the performance of greedy and non-adaptive schemes.

- **Random.** This algorithm selects m random beams, based on a uniform random distribution, in each period.

Performance metrics. The performance metrics used in the evaluation are aggregate and cumulative received (rx) data, the number of served vehicles, and average learning time (i.e., the time required for FML to reach a certain percentage of the Optimal's performance). The aggregate received data is defined as the data received (in bits) by all the vehicles in the system in time period t . The cumulative received data is defined as the data received by all the vehicles in the system from beginning of the simulation up to time period t .

C. Numerical Evaluation

Here, we first evaluate a generic scenario. Next, we analyze the impact of several parameters, i.e., the arrival rate of the vehicles, the number of selected beams, the frequency of blockages, the underlying traffic patterns, the average contact time, the adaptability, the radio performance versus environment, and performance comparison to a tracking mechanism. Unless otherwise stated, we consider the case where (i) the percentage of permanent and temporary blockages each corresponds to 20% of all paths and (ii) traffic patterns (i.e., the vehicle arrival rate and the route probabilities) change based on the typical patterns provided by Google Maps. Since Google Maps only provides the typical daily traffic patterns for 17 hours of the day (from 06:00 to 22:00), we run algorithms over a 17-hour simulation. These additional variables are introduced to evaluate the adaptability of FML to the environment dynamics (e.g., the location of temporary blockages, popular paths, channel qualities, and speed of the vehicles). Each simulation is repeated 20 times for which we show 95% confidence intervals in the figures.

1) **Average received data:** In the following, we analyze the aggregate received data achieved by the algorithms over a time horizon of 17 hours. Fig. 5 shows the aggregate received data per time period for an arrival rate of $\lambda = 0.4$ in the case of $m = 4$ selected beams per period. The fluctuations in the graph result from the number of vehicles in the system. Specifically, the aggregate received data increases with the number of

vehicles. The impact of vehicle arrival rate and traffic patterns are evaluated in the following sections in detail. As expected, Optimal gives an upper bound to the other algorithms due to a priori knowledge of the expected beam performance. Our proposed algorithm FML clearly outperforms the other algorithms UCB, MaxRate, and Random. We observe that FML's performance quickly approaches that of Optimal within the first 100 periods while the other algorithms perform at least $\sim 20\%$ worse than FML. This behavior is even more pronounced from the 256th period when FML remains very close to Optimal's aggregate received data. FML experiences small divergence (below 3%) from Optimal at a few points within the simulation. These small variations are due to: (i) the occurrence of new events, which are not learned from, or (ii) the re-exploration of past decisions, which is as expressed in (5). FML revisits past decisions to ensure that the historic performance values for a given decision is still valid. If the algorithm finds out that the observed performance of a decision is changed, it will update the learning accordingly. Note that at around time period 300, FML performs slightly better than Optimal. This occurrence is due to the fact that Optimal selects the beam that has the highest *expected* performance. However, in rare cases, the instantaneous performance of the beams selected by FML turns out to be higher. Average performance within 17 hours of simulation indicates that the average aggregate received data achieved by FML is 21.99%, 36.08%, and 54.76% higher than that achieved by UCB, MaxRate and Random, respectively. Moreover, on average, FML performs only 1.73% below that achieved by Optimal.

2) **Impact of arrival rate:** Next, we investigate the impact of the arrival rate on the cumulative received data achieved by the different algorithms in the case of $m = 4$ selected beams per period for different arrival rates $\lambda \in \{0.2, 0.4, 0.6, 0.8, 1\}$. From Fig. 6, we can observe that the cumulative received data grows as the number of vehicles in the system increases. Over the whole range of λ , the cumulative received data achieved by FML lies between 9.36% and 23.06% higher than that achieved by the next-best algorithm UCB and only up to 3.06% lower than that achieved by the Optimal.

Fig. 7 shows the time required for FML to achieve $\{80\%, 85\%, 90\%\}$ of the Optimal's performance (i.e., the average received data) for different λ , respectively. More specifically, we track the performance of our proposal against Optimal during the simulation and record the time at which FML achieve $\{80\%, 85\%, 90\%\}$ of the Optimal. It is observed from our evaluation that FML achieves $\{80\%, 85\%, 90\%\}$ of Optimal's performance within $\{13, 25, 56\}$ mins for all arrival rates, respectively. The large confidence interval is due to the randomness of certain parameters both in the learning algorithm (exploration decisions) and in the evaluation scenario (e.g., location of temporary blockages, selected routes, speed). As a result, FML may approach near-optimal performance below seven minutes if all random effects are in favor, and up to 75 minutes otherwise. Note that even manual configuration and war-driving tests require much more than 75 mins. Furthermore,

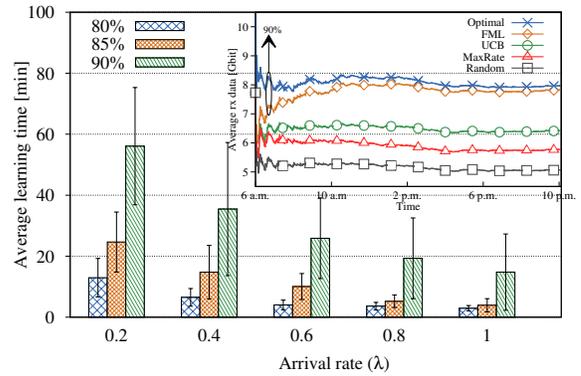


Fig. 7. Impact of arrival rate λ on required time for FML to achieve a certain percentage of Optimal's performance for $m = 4$.

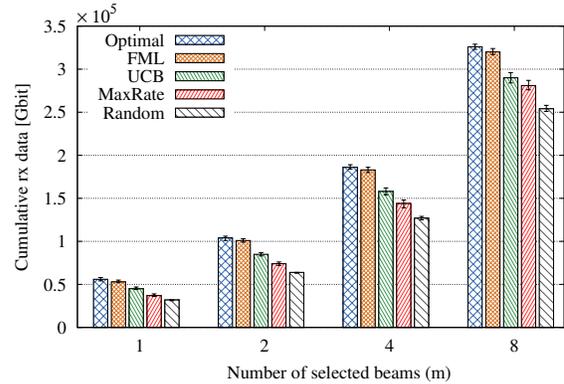


Fig. 8. Impact of number m of selected beams per period on cumulative received data for arrival rate $\lambda = 0.4$ after 17 hours.

war-driving tests may only capture effects of permanent, but not from temporary blockages. The convergence rate of a learning algorithm depends on the selection actions and the corresponding reward it observe, it thus varies tremendously. The subplot in Fig. 7 shows the average received (rx) data with $\lambda = 0.4$. Average rx data is the average data over all the vehicles in the system up to this time period. This figure illustrates FML's quick learning and adaptation capability. In particular, FML achieves 90% of the performance Optimal within 30 time periods. This result tallies with the performance figure shown for $\lambda = 0.4$ in Fig. 7 and the trend observed in Fig. 5. This shows how quickly FML converges to near-optimal beam selection. Moreover, the general trend shows that the time to converge to near-optimal results reduces when the vehicle density in the system increases. This is due to the fact that with higher vehicle density, FML has more examples to learn from simultaneously.

3) **Impact of the number of selected beams:** Here, we analyze the impact of the number of selected beams m per period on the cumulative received data. Fig. 8 shows the cumulative received data achieved with an arrival rate of $\lambda = 0.4$ for different $m \in \{1, 2, 4, 8\}$. As the number of simultaneously selected beams increases, the cumulative received data increases as well. This increase is due to the enhanced coverage area. However, as mentioned earlier, the higher the number of beams, the higher is the hardware

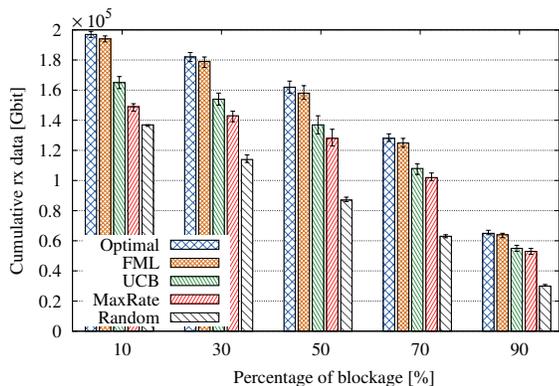


Fig. 9. Impact of blockages on cumulative received data for arrival rate $\lambda = 0.4$ after 17 hours.

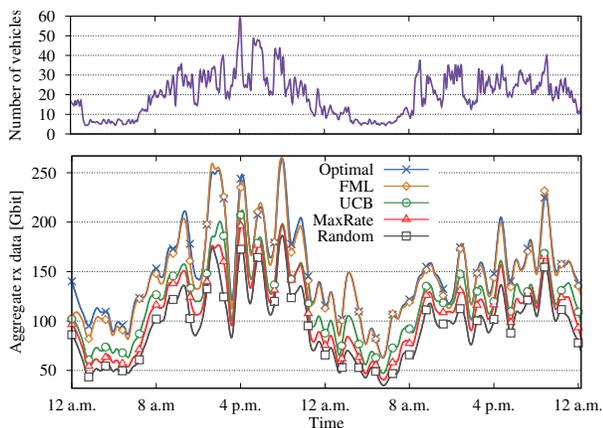


Fig. 10. Data for 48 hours of live daily traffic pattern.

complexity and energy consumption at the mmBS [5]. For different values of m , the cumulative received data achieved by FML is between 10.45% and 18.98% higher than that achieved by the next-best algorithm UCB and only up to 4.71% lower than that achieved by the Optimal.

4) **Impact of blockages:** Here, we investigate the impact of blockages on the cumulative received data. Fig. 9 shows the cumulative received data with an arrival rate of $\lambda = 0.4$ in case of $m = 4$ selected beams per period for $\{10\%, 30\%, 50\%, 70\%, 90\%\}$ of permanent blockages in the system. Clearly, as the percentage of permanent blockages in the system increases, the cumulative received data decreases². For any percentage of permanent blockages, FML outperforms all non-optimal algorithms. The cumulative received data achieved by FML lies between 15.55% and 17.42% higher than that achieved by the next-best algorithm, i.e., UCB. Moreover, FML's achieved results deviate from that of Optimal merely by at most 2.61%.

5) **Live daily traffic pattern:** The prior evaluation was based on the typical traffic pattern as in Fig. 4. Due to the averaging effect, Google's typical traffic does not capture the quick changes of traffic patterns which are visible in the live traffic report. To this aim, we recorded the observed live traffic reports of Google for a period of 48 hours in 30-minute

²We observed the same trend with temporary blockages. Due to the lack of space, the aforementioned results are not graphically demonstrated.

intervals. We fed this data to the simulator to evaluate the performance of FML in live traffic conditions. For instance, the arrival rate is higher during office hours and lower at other times of the day. Moreover, route probabilities also change according to the live statistics. The top plot of Fig. 10 shows the number of vehicles in the system within a 48-hour-period. Clearly, the arrival rate has characteristic peaks during the course of a day (especially between 6 a.m. and 10 p.m.), which lead to an increase in vehicle density. For better readability, the graph in Fig. 10 is smoothed by 5% and 1% for the figures at the top and bottom, respectively, with a local regression using weighted linear least squares and a 2nd degree polynomial model. The bottom plot of Fig. 10 shows the aggregate received data for $m = 4$ selected beams per period. We can see that FML achieves near-optimal aggregate received data within at most four hours. We also observe that FML can capture the effect of traffic fluctuations and leverage it to make better decisions. Since the other algorithms do not adapt to the change in traffic, they perform worse than FML. Averaging over 48 hours of simulation, FML performs 24.96%, 39.61%, and 60.51% better than UCB, MaxRate, and Random, respectively. Further, the performance of FML only lies within 2.47% below that of Optimal.

6) **Average contact time:** The beam selection decision in a dynamic environment (e.g., variable traffic pattern) is impacted by several factors. These factors include: (i) the time each vehicle remains within the beam coverage area (i.e., contact time) and (ii) the achievable rate by each beam. Intuitively, for the same vehicular's speed, a wider beam should result in higher contact time, as they provide a larger coverage area. However, in urban areas, the contact time is also influenced by the road topology and the speed of the vehicle. As mentioned in the simulation setup, the speed of the vehicles in our scenarios varies between 50km/h and 70 km/h.

Fig. 11 demonstrates the average contact time of each beam under different algorithms. As observed, the contact times are relatively similar regardless of the algorithm. This observation stems from the fact that the decision of an algorithm has no influence on the speed of the vehicles and road topology but vice-versa. In fact, the slight variation is just due to the variability in the vehicle speed. We also observe that the contact time of some beams is zero. Zero contact time implies either that a specific beam is never selected by a given algorithm or when a beam is selected, the path that is covered by it, is not traversed by any vehicle. Therefore, no statistical data is available for that beam. Fig. 11 also shows that contact time is not always dependent on the beam width, but the distribution of speed of the vehicles that pass by the beam. For example, beam 7 and 15 have similar beam width ($\sim 11^\circ$), but the average contact of beam 7 is significantly less than that of 15.

7) **Adaptability:** Here we take a closer look at the selected beams by different algorithms in order to determine their adaptability. In particular, we leverage pie charts to show the frequency of each beam being selected. This form of result representation assists us to visualize the capability of each

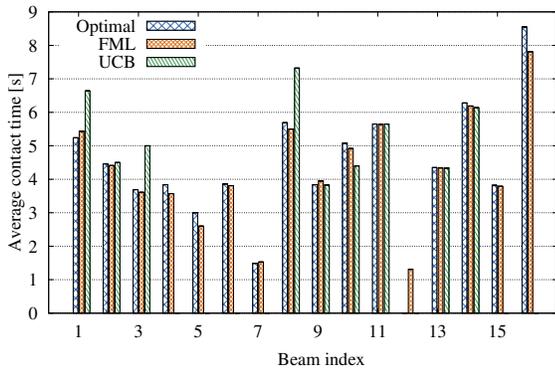


Fig. 11. Average contact time of each beam under different algorithms.

algorithm to adapt to environmental changes by exploiting the available beam patterns; this can be concluded by comparing the selected beam of each algorithm to that of `Optimal`. Furthermore, we can analyze the similarity of the decision and strategy taken by the algorithms.

In Fig. 12, we can observe the portion of time each beam is used by the `Optimal`, `FML`, and `UCB` algorithm for the different number of possible orthogonal beams, i.e., $m \in \{1, 2, 4, 8\}$. Our first observation is that; on the one hand, regardless of the values of m , `FML`'s beam selection is very similar to that of `Optimal`. This tallies with our prior observation (i.e., its performance on aggregate and cumulative received data) and shows that `FML` adapts almost optimally to the traffic dynamics. On the other hand, `UCB` confines its selection to a small subset of beams as it does not take into account the context information obtained from the environment. Furthermore, we can see that the set of selected beams extends as m increases. It is interesting to note that the beams selected in $m \in \{1, 2\}$ (such as 11 and 15) will still be dominant for higher m . In addition, it can be observed that beams 2, 8, 11, and 15 are the preferred choices. We will take a closer look and further analyze the reasons behind such beam selection policies in the following subsection.

8) **Radio performance vs. environment:** As observed above, the widest beam is not necessarily the best beam. Here, we discuss the impact of other factors such as the length of roads which fall within the coverage of a specific beam (i.e., covered path) and the achievable data rate, which itself depends on the received signal quality. The achievable rate depends on the beam width and the distance of a point on the road to the mmBS. To this aim we analyze Fig. 13(a) and Fig. 13(b) side-by-side. On the one hand, Fig. 13(a) shows a plot of the covered path and the respective beam width of each beam. On the other hand, Fig. 13(b) depicts the covered path with respect to the average achievable rate when a beam is used. The average rate is obtained by averaging the rate at all the possible points on the road that is covered by a particular beam.

It is observed in Fig. 13(a) that a wider beam does not necessarily cover a longer stretch of the road. This behavior is particularly obvious for beams 5, 6, 12, 13, and 14, which is

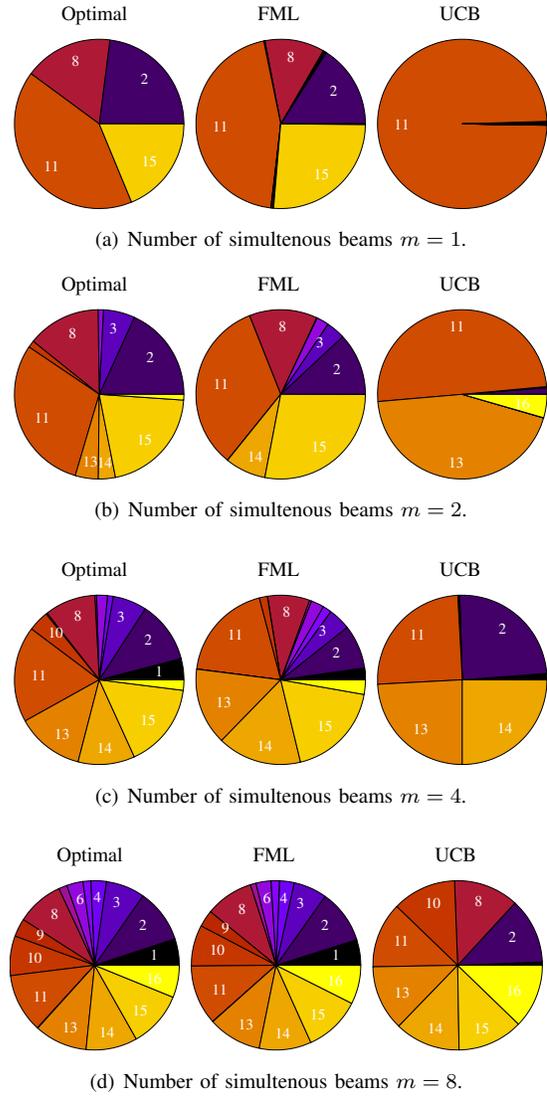
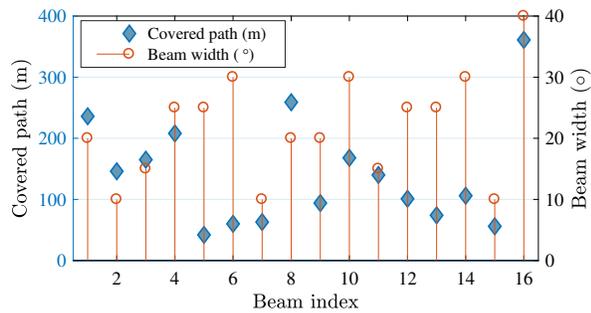


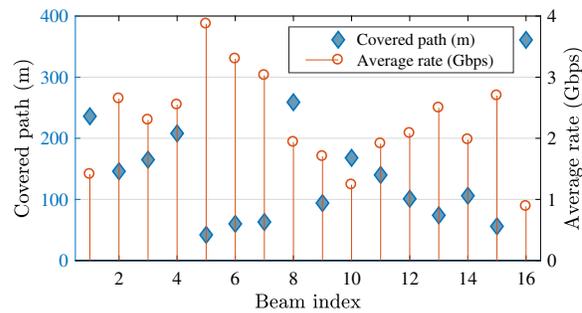
Fig. 12. Fraction of time a beam is selected by `Optimal`, `FML`, and `UCB`. The beam selection of `FML` is clearly very similar to that of `Optimal`. This shows that `FML` adapts almost optimally to the traffic dynamics.

caused by the topology of the road. For example, a narrower beam that is pointing in the direction of an intersection potentially covers a larger portion of the road, while a wider beam that points towards a straight road may not have a long covered path.

Fig. 13(b) visualizes the achievable data rate for each beam and its association with the length of the covered path. As illustrated, beams 5, 6, and 7 provide the highest data rates. However, referring to the pie charts in Section V-C7, none of these beams are among the dominant selected beams. This is due two main reasons: (i) the path which contributes to a high average rate is less traversed by the vehicle, (ii) although the average rate achieved by a beam is high, the total rate is highly dependent on the rate achieved throughout the path, which may be low. More specifically, the beam with the highest data rate is not necessarily the best if it provides a very short contact time due to street topology and the traffic pattern. For these



(a) Covered path versus beam width.



(b) Covered path versus average data rate.

Fig. 13. Information on the covered path, beam width and average data rate for each beam.

reasons, selecting those beams on the basis of data rate can be indeed detrimental to the overall performance.

Analyzing the above-mentioned results side-by-side helps us to verify that the dominant beams (i.e., 2, 8, 11, and 15) are chosen because they provide a good trade-off between the average rate and the covered path. The last ambiguity is on the selection of beam 15 over beams 6 and 7. We observe that the rate-coverage trade-off of 15 is somewhat similar to that of 6 and 7, but it is still selected more often. As explained, the reason behind this preference is the dominant traffic direction. Specifically, beam 15 covers a more crowded stretch of the road. As a result, both Optimal and FML choose beam 15 over 13 and 14. However, UCB's chose beam 13 and 14 more frequently since it does not have the notion of dominant traffic (see Fig. 12). This result further emphasizes the importance of traffic awareness in mmWave V2X scenarios.

9) **To track or not to track:** Tracking individual vehicles is a common approach in the majority of the literature [22]. From the implementation point of view, this approach is very challenging because the network needs to constantly acquire/predict the vehicle's location and trajectory. From a network capacity point of view, tracking individual vehicles may result in a very low number of served vehicles, specifically in dense networks [23]. This is in particular important if the number of simultaneous beams is limited due to hardware limitation. In this scenario, we highlight the impact of providing selective coverage areas to all vehicles versus tracking individual vehicles. To illustrate this, we additionally implemented an optimal tracking (i.e., OptTrack) algorithm, which has an instantaneous localization knowledge of the vehicle. Fig. 14(a) shows the cumulative received data and the average number of served vehicles per period achieved by FML and OptTrack for $m = 4$ after 17 hours for different arrival rates λ . For a fair comparison, we assume OptTrack may track as many vehicles simultaneously, as FML may select beam patterns simultaneously. We observe that OptTrack has a good performance as long as the number of vehicles in the system is very low. This is because OptTrack follows a few vehicles perfectly, which in this case is better than sticking to a fixed subset of beams over the duration of a period as done by FML. However, as the arrival rate and the number of vehicles in the system increases, FML outperforms

OptTrack both in terms of the cumulative received data (up to 61.37%) and the average number of served vehicles (up to 82.55% vehicles more per period). FML achieves higher gains for higher arrival rate because it provides larger and better coverage in areas with higher traffic. In addition, FML serves each vehicle passing through the selected beam, while OptTrack only serves a limited set of perfectly tracked vehicles while they reside within the coverage area of the mmBS.

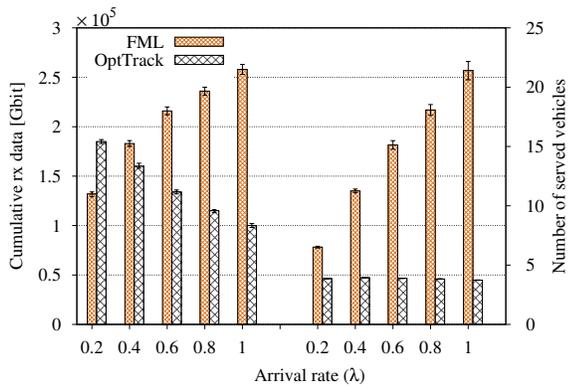
In Fig. 14(b), we observe the fraction of time each beam is selected by OptTrack for a different number of simultaneous beams $m \in \{1, 2, 4\}$. In this scenario, once a vehicle is served by the mmBS, it will be tracked until it leaves the system. Given that OptTrack's beam selection simply depends on the chosen path of the vehicles and other parameters (e.g., coverage area, data rate) do not play a role, we can observe that the set of selected beams by OptTrack is rather uniform. Specifically, OptTrack neglects opportunistic gain, and it has no prioritizing mechanism among different beams.

VI. DISCUSSION

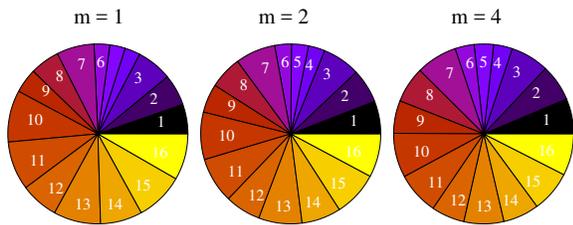
Our evaluation results confirm the superiority and adaptability of FML compared to the benchmark algorithms. In this section, we discuss the impact of some of our system model assumptions (e.g., beam orthogonality) and the capabilities of FML, which were not detailed previously.

A. Beam orthogonality

We assumed that the available beams at the mmBS are all orthogonal. Nevertheless, FML's performance is not restricted to this limitation. In fact, non-orthogonality can be formulated as an additional constraint in our model, where overlapping beams cannot be used simultaneously. In this case, the set of actions (i.e., different *beam combinations*) naturally increases. However, not all possible combinations are allowed due to the constraint that overlapping beams may not be used simultaneously. Thus, the number of actions remains feasible. Nevertheless, the convergence time of the FML (similar to any other learning-based approach) depends on the size of the available actions. Using online learning algorithms (such as FML) becomes necessary in such cases to ensure the system remains operational before the algorithm converges to (near)-optimal solution.



(a) Impact of number m of selected beams per period on cumulative received data for arrival rate $\lambda = 0.4$ after 17 hours.



(b) Fraction of time a beam is selected by OptTrack.

Fig. 14. Impact of individual vehicle tracking using OptTrack on the network performance.

B. Number of selected beams

We assumed that the number m of simultaneously selectable beams at the mmBS is limited due to current hardware limitation [5]. Assuming fully digital beamforming materializes in the future, this limitation will present itself as overlapping beam patterns, which cannot be selected simultaneously. FML can also be used in such scenarios to select the most suitable beam pattern according to the available context (e.g., vehicle's location).

C. Interference

In this paper, we assume that there is no interference among orthogonal beams. However, *theoretically* orthogonal beams may interfere with each other due to reflections from surrounding objects (e.g., walls, buildings). Such scenarios are very rare in high frequency ranges of mmWave bands such as 60 GHz due to low reflection coefficient but more plausible in 28 GHz and 38 GHz. Although this factor is not considered in our simulation, its impact will be interpreted as a blockage by FML, thus refraining from communicating over those beams.

D. Co-located vehicles

There could be multiple vehicles in the coverage area of a beam at the same time. Given the focus of this paper is on unicast communication, only one of these vehicles is allowed to communicate with the mmBS. In our simulation, we select one of the vehicles at random for communication. While scheduling is out of our scope, the random selection can be easily adapted to a scheme which takes into account the past

throughput of the co-located vehicles to increase the fairness. More importantly, it should be noted that this phenomenon has a high potential for multicast vehicular scenarios.

E. Location reporting

FML infers traffic patterns from a very coarse location information (i.e., direction of arrival). This level of coarseness is intentionally chosen to emphasize on the potential of FML to infer traffic patterns based on coarse geo-locations. If the network can afford the extra overhead for high-resolution location reporting, FML's performance will only improve further.

F. FML for tracking

We have shown in the evaluation that tracking individual vehicles is less efficient. Our system model is designed to provide selective coverage areas since V2X mmWave systems are motivated by the need to exchange intermittent large data files [2], [3]. Nevertheless, FML can be adapted to track individual vehicles by using context such as speed of the vehicle, traffic patterns, and blockages to estimate the vehicle's future location and to select the suitable beam width accordingly.

VII. RELATED WORK

Beam selection issues have been addressed before in conventional vehicular and cellular networks operating at sub-6 GHz frequencies to achieve maximum rate using multi-lobe beam patterns. Unlike our proposal, these works rely on accurate GPS location reporting in order to perform beam switching. The complexity of this method grows exponentially with variable vehicular speed and channel conditions. In addition, the proposed algorithms are not able to adapt to environmental change such as that considered in this manuscript. Most importantly, as mentioned, the signal propagation characteristics in sub-6 GHz frequencies fundamentally differs from that of mmWave bands.

This paper proposes an adaptable learning algorithm for mmWave vehicular scenarios in which blockages and traffic are taken into account. Specifically, our algorithm does not require either accurate localization information or prior statistical knowledge of the variability or change in the traffic and environment. Therefore, its performance is independent of the aforementioned variabilities. In [24], MAB learning algorithm is used for maximizing the directivity gain through efficient beam alignment between mmWave transceivers. This work suits applications which require tracking to prolong the contact time between the transceivers. In essence, this work could be used to extend FML to achieve longer connectivity, at the cost of higher complexity and additional constraints based on the system's objective function. The increased complexity stems from the additional learning required for beam alignment. In what follows, we provide a short overview of the ongoing efforts in mmWave vehicular research. For more details, we encourage the readers to refer to a recent survey on the topic in [7]. The body of the works on mmWave V2X can be categorized in channel characterization, PHY design, and MAC design.

In [25], Kato et al. provide propagation characteristics of mmWave communication in inter-vehicular scenarios. On the other hand, the authors of [26] derive closed-form approximation for coherence time and beam width with consideration of directional communication. The feasibility of mmWave communication is analyzed in [4] via an extensive measurement campaign. The results indicate the low number of scattering cluster in mmWave bands, which imply that the number of supported data streams are significantly less than the antenna array size. The challenges of enabling mmWave V2X communication is elaborated in [27]. In addition, the authors describe possible solutions to these challenges from PHY and MAC perspective. The works in [28]–[31] focus on mmWave beam adaptation in vehicular scenarios. These works exploit dedicated short-range communications (DSRC) to estimate the location of the vehicles. This estimation allows mmBS to track the moving vehicles and to adapt the mmWave beam accordingly. While feasible, this technique requires a complex transceiver chain and accurate localization information. Further, modification to the MAC protocol to allow the exchange of this information between the interfaces is crucial.

While prior work agrees that blockage is the Achilles heel of mmWave communication [4]–[7], [27], most work focus on modeling the blockages. Tassi *et al.* use a stochastic geometry tool to derive an approximation for the achievable signal quality. In [32], Wang *et al.* also take blockages into account and model the closed-form expression using Manhattan Poisson Line Process street model. These works provide highly relevant insight to the performance bound of mmWave communication in a vehicular network. However, they do not provide a method to automatically detect the change in the state of the network (i.e., blockages and traffic change) which allow it to adapt accordingly. Furthermore, traffic-awareness is another aspect which has not been addressed to the best of our knowledge.

VIII. CONCLUSIONS

In this paper, we address the problem of beam selection at mmWave base stations where the outcome of the selection is highly dependent on the traffic and the blockages in the network. To this aim, we propose FML, an online learning algorithm based on contextual multi-armed bandits that operates on minimal contextual network information (i.e., a vehicle's direction of arrival). In addition, we analyze the implementation feasibility of FML in the cellular network by proposing a protocol within the definition of 3GPP standard. The advantage of FML is twofold: (i) it enables mmWave base stations to autonomously learn from the context to understand their surrounding environment and (ii) it provides a scalable solution to increase the deployment density of mmWave base stations with minimal setup overhead for the operators. Our evaluation results show that FML requires on average only 33 mins to achieve near-optimal performance. Noteworthy, without the overhead of tracking of OptTrack, FML achieves 61.37% and 82.55% gain in terms of the cumulative received data and number of served vehicles, respectively. The results

demonstrate the capability of online bandit learning and emphasize on the relevance of context-awareness in 5G scenarios. Furthermore, exploring a hybrid solution between tracking individual vehicles and increasing the overall network capacity appears to be an interesting future research avenue.

This is the first attempt to incorporate learning algorithms in such dynamic vehicular scenarios. However, our analytical modeling of the system can be extended to observe and learn from richer context (e.g., coordinates and type of the vehicle). Given the low complexity of contextual bandit learning, this approach can be applied in other areas of mmWave cellular networking such as initial access and user handover.

IX. ACKNOWLEDGEMENTS

This work has been co-funded by the LOEWE initiative (Hessen, Germany) within the NICER project and the German Research Foundation (DFG) in the Collaborative Research Center (CRC) 1053 - MAKI.

REFERENCES

- [1] A. Asadi, G. Sim, S. Müller, A. Klein, and M. Hollick, "FML: Fast Machine Learning for 5G mmWave Vehicular Communications," in *IEEE INFOCOM*, 2018.
- [2] K. Sakaguchi, T. Haustein, S. Barbarossa, E. C. Strinati, A. Clemente, G. Destino, A. Pärssinen, I. Kim, H. Chung, J. Kim, W. Keusgen, R. J. Weiler, K. Takinami, E. Ceci, A. Sadri, L. Xain, A. Maltsev, G. K. Tran, H. Ogawa, K. Mahler, and R. W. Heath Jr, "Where, When, and How mmWave is Used in 5G and Beyond," *arXiv.org*, Apr 2017.
- [3] J. Choi, V. Va, N. Gonzalez-Prelcic, R. Daniels, C. R. Bhat, and R. W. Heath, "Millimeter-Wave Vehicular Communication to Support Massive Automotive Sensing," *IEEE Communications Magazine*, vol. 54, no. 12, pp. 160–167, Dec 2016.
- [4] T. S. Rappaport, S. Sun, R. Mayzus, H. Zhao, Y. Azar, K. Wang, G. N. Wong, J. K. Schulz, M. Samimi, and F. Gutierrez, "Millimeter Wave Mobile Communications for 5G Cellular: It Will Work!" *IEEE Access*, vol. 1, pp. 335–349, May 2013.
- [5] S. Rangan, T. S. Rappaport, and E. Erkip, "Millimeter-wave Cellular Wireless Networks: Potentials and Challenges," in *Proceedings of the IEEE*, Feb 2014.
- [6] A. Alkhateeb, Y.-H. Nam, M. S. Rahman, J. Zhang, and R. W. Heath, "Initial Beam Association in Millimeter Wave Cellular Systems: Analysis and Design Insights," *IEEE Transactions on Wireless Communications*, vol. 16, no. 5, pp. 2807–2821, Mar 2017.
- [7] V. Va, T. Shimizu, G. Bansal, and R. W. Heath Jr, "Millimeter Wave Vehicular Communications: A Survey," *Foundations and Trends in Networking*, vol. 10, no. 1, pp. 1–113, Jun 2016.
- [8] L. Wei, R. Q. Hu, Y. Qian, and G. Wu, "Key Elements to Enable Millimeter Wave Communications for 5G Wireless Systems," *IEEE Wireless Communications*, vol. 21, no. 6, pp. 136–143, Dec 2014.
- [9] M. Series, "Guidelines for Evaluation of Radio Interface Technologies for IMT-Advanced," *Report ITU*, no. 2135-1, Dec 2009.
- [10] S. Han, C. I. I, Z. Xu, and C. Rowell, "Large-scale Antenna Systems with Hybrid Analog and Digital Beamforming for Millimeter Wave 5G," *IEEE Communications Magazine*, vol. 53, no. 1, pp. 186–194, Jan 2015.
- [11] H. Shokri-Ghadikolaei, F. Boccardi, E. Erkip, C. Fischione, G. Fodor, M. Kountouris, P. Popovski, and M. Zorzi, "The Impact of Beamforming and Coordination on Spectrum Pooling in MmWave Cellular Networks," in *ASIOMAR*, Nov 2016.
- [12] M. Giordani, M. Mezzavilla, and S. Rangan, "Multi-Connectivity in 5G mmwave cellular networks," in *IFIP Med-Hoc-Net*, Jun 2016.
- [13] H. Shokri-Ghadikolaei, F. Boccardi, C. Fischione, G. Fodor, and M. Zorzi, "Spectrum Sharing in mmWave Cellular Networks via Cell Association, Coordination, and Beamforming," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 11, pp. 2902–2917, Nov 2016.
- [14] S. Maghsudi and E. Hossain, "Multi-armed Bandits with Application to 5G Small cells," *IEEE Wireless Communications*, vol. 23, no. 3, pp. 64–73, Jun 2016.

- [15] P. Auer, N. Cesa-Bianchi, and P. Fischer, "Finite-time Analysis of the Multiarmed Bandit Problem," *Journal of Machine Learning*, vol. 47, no. 2-3, pp. 235–256, May 2002.
- [16] T. Lu, D. Pal, and M. Pal, "Contextual Multi-Armed Bandits," in *AISTATS*, May 2010.
- [17] A. Slivkins, "Contextual Bandits with Similarity Information," *Journal of Machine Learning Research*, vol. 15, no. 1, pp. 2533–2568, Jan 2014.
- [18] C. Tekin, S. Zhang, and M. van der Schaar, "Distributed Online Learning in Social Recommender Systems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 4, pp. 638–652, Aug 2014.
- [19] S. Müller, O. Atan, M. van der Schaar, and A. Klein, "Context-Aware Proactive Content Caching with Service Differentiation in Wireless Networks," *IEEE Transactions on Wireless Communications*, vol. 16, no. 2, pp. 1024–1036, Feb 2017.
- [20] 3GPP, "Technical Specification Group Radio Access Network: Study on channel model for frequencies from 0.5 to 100 GHz," Tech. Rep., Mar 2017.
- [21] T. Nitsche, G. Bielsa, I. Tejado, A. Loch, and J. Widmer, "Boon and Bane of 60 GHz Networks: Practical Insights into Beamforming, Interference, and Frame Level Operation," in *ACM CoNEXT*, May 2015.
- [22] V. Va, H. Vikalo, and R. W. Heath, "Beam Tracking for Mobile Millimeter Wave Communication Systems," in *IEEE GlobSIP*, Dec 2016, pp. 743–747.
- [23] A. Loch, A. Asadi, G. H. Sim, J. Widmer, and M. Hollick, "mm-Wave on Wheels: Practical 60 GHz Vehicular Communication without Beam Training," in *COMSNETS*, Jan 2017, pp. 1–8.
- [24] M. Hashemi, A. Sabharwal, C. E. Koksal, and N. B. Shroff, "Efficient Beam Alignment in Millimeter Wave Systems Using Contextual Bandits," *CoRR*, vol. abs/1712.00702, 2017. [Online]. Available: <http://arxiv.org/abs/1712.00702>
- [25] A. Kato, K. Sato, and M. Fujise, "ITS Wireless Transmission Technology. Technologies of Millimeter-wave Inter-vehicle Communications: Propagation Characteristics," *Journal of the Communications Research Laboratory*, vol. 48, no. 1, pp. 99–110, Nov 2001.
- [26] V. Va and R. W. Heath, "Basic Relationship Between Channel Coherence Time and Beamwidth in Vehicular Channels," in *IEEE VTC*, Jan 2015.
- [27] M. Giordani, A. Zanella, and M. Zorzi, "Millimeter Wave Communication in Vehicular Networks: Challenges and Opportunities," in *IEEE MOCAS*, May 2017.
- [28] V. Va, T. Shimizu, G. Bansal, and R. W. Heath, "Beam Design for Beam Switching Based Millimeter Wave Vehicle-to-infrastructure Communications," in *IEEE ICC*, Jul 2016.
- [29] J. Choi, V. Va, N. Gonzalez-Prelcic, R. Daniels, C. R. Bhat, and R. W. Heath, "Millimeter-Wave Vehicular Communication to Support Massive Automotive Sensing," *IEEE Communications Magazine*, vol. 54, no. 12, pp. 160–167, Dec 2016.
- [30] I. Mavromatis, A. Tassi, R. J. Piechocki, and A. R. Nix, "MmWave System for Future ITS: A MAC-layer Approach for V2X Beam Steering," *CoRR*, vol. abs/1705.08684, May 2017. [Online]. Available: <http://arxiv.org/abs/1705.08684>
- [31] —, "Beam Alignment for Millimeter Wave Links with Motion Prediction of Autonomous Vehicles," *CoRR*, vol. abs/1702.04264, Feb 2017. [Online]. Available: <http://arxiv.org/abs/1702.04264>
- [32] Y. Wang, K. Venugopal, A. F. Molisch, and R. W. Heath, "Blockage and Coverage Analysis with MmWave Cross Street BSs Near Urban Intersections," pp. 1–6, May 2017.



Dr. Gek Hong (Allyson) Sim is currently a Post-Doctoral researcher at the Department of Computer Science, Technische Universität Darmstadt, Germany. Allyson received her Bachelor Degree (Engineering majoring in Telecommunication) and first Master Degree (Engineering Science) from Multimedia University, Malaysia in 2007 and 2011, respectively. She was awarded with Master Degree and PhD Degrees in Telematics Engineering from University Carlos III Madrid in 2012 and 2015, respectively. Her research interest are multicast scheduling and MAC layer optimization for millimeter-wave networks.

scheduling and MAC layer optimization for millimeter-wave networks.



Sabrina Klos (née Müller) (S'15) received the B.Sc. and M.Sc. degrees in mathematics from Technische Universität Darmstadt, Germany, in 2012 and 2014, respectively, where she is currently pursuing the Ph.D. degree in electrical engineering as a member of the Communications Engineering Laboratory. Her research interests include machine learning and optimization methods and their applications to wireless networks.



IEEE COMSOC Tech Focus.

Dr. Arash Asadi is appointed as Athena Young Investigator of Technische Universität Darmstadt. His research interests include several aspects of wireless communication: 5G cellular networking, millimeter-wave communication, Device-to-Device (D2D) communication techniques, Software-defined Radio (SDR) prototyping. He is a recipient of several awards including outstanding PhD and master thesis awards from UC3M. Some of his papers on D2D communication has appeared in IEEE COMSOC best reading topics on D2D communication and in



IEEE COMSOC Tech Focus.

Prof. Dr.-Ing. Anja Klein (M'96) received the Diploma and Dr.-Ing. (Ph.D.) degrees in electrical engineering from the University of Kaiserslautern, Germany, in 1991 and 1996, respectively. In 1996, she joined Siemens AG, Mobile Networks Division, Munich and Berlin. She was active in the standardization of third generation mobile radio in ETSI and in 3GPP, for instance leading the 3GPP RAN1 TDD group. She was director of a development department and a systems engineering department. In 2004, she joined the Technische Universität Darmstadt, Germany, as full professor, heading the Communications Engineering Laboratory. Her main research interests are in mobile radio, including interference management, cross-layer design, relaying and multi-hop, computation offloading, smart caching and energy harvesting. Dr. Klein has authored over 280 peer-reviewed papers and has contributed to 12 books. She is inventor and co-inventor of more than 45 patents in the field of mobile communications. In 1999, she was named the Inventor of the Year by Siemens AG. She is a member of Verband Deutscher Elektrotechniker - Informationstechnische Gesellschaft (VDE-ITG).



Prof. Dr.-Ing. Matthias Hollick is currently heading the Secure Mobile Networking Lab in the Computer Science Department of Technische Universität Darmstadt, Germany. After receiving the Ph.D. degree from TU Darmstadt in 2004, he has been researching and teaching at TU Darmstadt, Universidad Carlos III de Madrid, and the University of Illinois at Urbana-Champaign. His research focus is on resilient, secure, privacy-preserving, and quality-of-service-aware communication for mobile and wireless systems and networks.