Adaptive Millimeter-Wave Communications Exploiting Mobility and Blockage Dynamics

Muddassar Hussain[†], Maria Scalabrin[‡], Michele Rossi[‡], and Nicolò Michelusi[†]

***Abstract*—Mobility may degrade the performance of next-generation vehicular networks operating at the millimeter-wave spectrum: frequent loss of alignment and blockages require repeated beam training and handover, thus incurring huge overhead. In this paper, an adaptive and joint design of beam training, data transmission and handover is proposed, that exploits the mobility process of mobile users and the dynamics of blockages to optimally trade-off throughput and power consumption. At each time slot, the serving base station decides to perform either beam training, data communication, or handover when blockage is detected. The problem is cast as a partially observable Markov decision process, and solved via an approximate dynamic programming algorithm based on PERSEUS [2]. Numerical results show that the PERSEUS-based policy performs near-optimally, and achieves a 55% gain in spectral efficiency compared to a baseline scheme with periodic beam training. Inspired by its structure, an adaptive heuristic policy is proposed with low computational complexity and small performance degradation.**

I. Introduction

Millimeter-wave (mm-wave) is a leading candidate to support the high capacity demands of future vehicular communications [3]. However, communication at these frequencies relies on highly directional transmissions and it is highly susceptible to blockages and mis-alignment. These features are exacerbated in highly mobile environments, resulting in degraded system performance. To compensate for these effects, the key question addressed in this paper is the following: *How can we leverage the information on the system dynamics (mobility of users and blockage dynamics) to optimize the communication performance? How much do we gain by doing so?* To address these questions, we envision the use of adaptive communication strategies and their formulation via partially observable (PO) Markov decision processes (MDPs).

We consider two base stations (BSs) serving a mobile user (MU) on both sides of a road link. At any time, the MU is associated with one of the two BSs (the serving BS). To enable directional data transmission (DT), the serving BS performs beam training (BT); to compensate for blockage, it performs handover (HO) to the other BS on the opposite side of the road link. The goal is to design the BT/DT/HO strategy, so as to optimally trade-off the throughput delivered to the MU and the average power consumption of BS. We formulate the optimization problem as a POMDP, and develop an approximate dynamic programming algorithm based on PERSEUS [2]. Our numerical evaluations based on a Gauss-Markov mobility model demonstrate that the PERSEUS-based policy performs very closely to a genie-aided upper bound in which the position of the MU and the blockage states are known, and outperforms a baseline scheme with periodic beam training by up to 55% in spectral efficiency. Motivated by the structure of the PERSEUS-based policy, we design an adaptive heuristic policy with low computational cost, and show numerically that it incurs a small 10% degradation in spectral efficiency compared to the PERSEUS-based policy.

Related Work: In the past decade, the design of beam training schemes for mm-wave systems has been the focus of extensive research, ranging from beam sweeping [4], estimation of angles of arrival (AoA) and of departure (AoD) [5], to data-assisted schemes [6], and feedback-based schemes [7]. Despite their simplicity, the overhead incurred by these algorithms may ultimately offset the benefits of beamforming in highly mobile environments [3]. *In this paper, we contend that leveraging a priori information on the vehicle's mobility as well as blockage dynamics may greatly improve the performance in vehicular communications* [8]. To this end, in [4], we designed optimal beam-sweeping schemes based on a worst-case mobility pattern. In [9], we designed adaptive strategies for BT/DT that leverage a priori mobility information, but with no consideration of blockage, hence no handover. In this work, we exploit both mobility and blockage dynamics to design adaptive communications schemes via POMDP.

Related work that applies learning techniques to mm-wave networks includes [10]–[12], revealing a growing interest in the design of adaptive communication policies that exploit side information to enhance the overall network performance. For instance, *contextual information* is exploited in [10] to reduce the training overhead, and the feedback is used in [11] to improve the beam search in the next rounds. However, these works neglect the impact of realistic mobility and blockage processes on the performance. In [12], the serving BS predicts blockages using past observations, and proactively performs handover to another BS with highly probable LOS link. However, the MU speed is randomly selected from a predefined set of values, and thus does not follow a realistic mobility process. Compared to this line of works, in this paper, we design adaptive communication strategies that leverage statistical information on the mobility and blockage processes in the selection of BT/DT/HO actions, with the goal to optimize the average long-term communication performance of the system. This approach is in contrast to strategies that either lack a mechanism to perform handover [10], [11], or assume a non realistic mobility pattern in their design [12].

II. System Model

We consider the scenario depicted in Fig. 1, where two BSs on both sides of a road link serve a MU moving along it. At any time, the MU is associated with one BS, denoted as the *serving BS*, which performs data transmission (DT) to the MU using beamforming to create a directional link, along with

[‡]School of Electrical and Computer Engineering, Purdue University, email: {hussai13,michelus}@purdue.edu

[†]Dept. of Information Engineering, University of Padova, email: {scalabri, rossi}@dei.unipd.it

An extended version of this paper appears in [1].

This research has been funded in part by NSF under grant CNS-1642982.

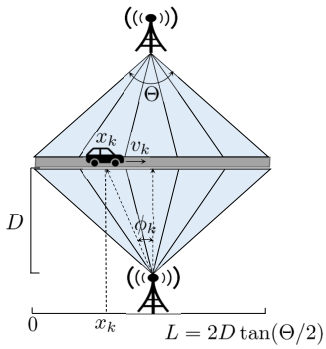

Fig. 1: A cell deployment with BSs on both side of road.

beam training (BT) to maintain alignment. The communication link between the serving BS and the MU is subject to time-varying blockage, which causes the signal quality to drop abruptly and DT to fail. To compensate for it, the serving BS may perform handover (HO) to the other BS on the opposite side of the road link, which then continues the process of BT and DT, until either another blockage event is detected, or the MU exits the coverage area of the two BSs. In this context, we investigate the design of the BT/DT/HO strategy, so as to optimize a trade-off between maximizing the throughput delivered to the MU and minimizing the power consumption of the BS during a transmission episode, defined as the time interval between the two instants when the MU enters and exits the coverage area of the two BSs. Both BSs are at a distance D from the road segment, symmetrically with respect to the road, and use a discrete set of narrow beams to communicate with the MU. To this end, the road segment covered by the two BSs, of length $L \triangleq 2D\tan(\Theta/2)$ and angular range Θ, is partitioned into S sectors of equal length $\Delta_\mathrm{s}=L/S$, indexed by $s\in\mathcal{S}\equiv\{1,\dots,S\}$. Each sector is then associated with one transmission beamformer $\mathbf{c}^{(s)}$, with angular support

$$\Phi_s=\left[\arctan\frac{(s-1)\Delta_\mathrm{s}-L/2}{D},\arctan\frac{s\Delta_\mathrm{s}-L/2}{D}\right],\ \forall s\in\mathcal{S},$$

and beamwidth $\theta_s=|\Phi_s|$, so that the ensemble of all beams span the entire angular region covered by the two BSs. $\mathbf{c}^{(s)}$ can be defined with a proper beam design, as done in the numerical results in Sec. V with the algorithm of [13].

Time is discretized into time-slots of duration Δ_t, corresponding to a beacon signal during BT or a data fragment during DT. Let $Z_k\in\bar{\mathcal{S}} \triangleq \mathcal{S}\cup\{\bar{s}\}$ denote the sector occupied by the MU at time k, where $Z_k=\bar{s}$ indicates that the MU exited the coverage area of the BSs. As a result of mobility of the MU, we model Z_k as a discrete-time Markov chain over $\bar{\mathcal{S}}$, with transition probabilities $\mathbf{P}_{ss'} = \mathbb{P}(Z_{k+1} = s'|Z_k = s)$. In the numerical results, we estimate $\mathbf{P}$ from time-series generated with the Gauss-Markov mobility model, in which the position x_k and speed v_k of the MU evolve as

$$v_k = \gamma v_{k-1} + (1-\gamma)\mu_v + \sigma_v\sqrt{1-\gamma^2}\tilde{v}_k, \quad (1)$$

$$x_k = x_{k-1} + \Delta_\mathrm{t} v_{k-1}, \quad (2)$$

where μ_v and σ_v are the average and standard deviation of v_k; γ is a memory parameter and $\tilde{v}_k\sim\mathcal{CN}(0,1)$, i.i.d. over k.

Within the kth time-slot of duration Δ_t, L symbols each of duration Δ_t/L are transmitted by the serving BS, denoted by the index $I_k \in \{1,2\}$. Let $\mathbf{x}_k \in \mathbb{C}^L$ be the signal transmitted such that $\mathbb{E}[\|\mathbf{x}_k\|_2^2] = L$. Assuming isotropic reception at the MU, the received signal is expressed as

$$\mathbf{y}_k = \sqrt{P_k}\mathbf{h}_k\mathbf{c}_k\mathbf{x}_k + \mathbf{w}_k, \quad (3)$$

where P_k is the transmit power of the serving BS; $\mathbf{h}_k\in\mathbb{C}^{1\times M_\mathrm{tx}}$ is the channel vector; M_tx is the number of antenna elements at each BS; $\mathbf{c}_k\in\mathbb{C}^{M_\mathrm{tx}\times 1}$ with $\|\mathbf{c}_k\|_2^2= 1$ is the beamforming vector; $\mathbf{w}_k\sim\mathcal{CN}(0,\sigma_w^2\mathbf{I})$ with $\sigma_w^2 = N_0 W_\mathrm{tot}$ is additive white Gaussian noise (AWGN), N_0 is the noise power spectral density, W_tot is the signal bandwidth.

In this paper, we model the channel as a single LOS path with binary blockage state $b_k^{(i)}\in\{0,1\}$ [14],

$$\mathbf{h}_k = \sqrt{M_\mathrm{tx}}b_k^{(I_k)}h_k\mathbf{d}_\mathrm{tx}(\psi_k)^H, \quad (4)$$

where $b_k^{(i)}=1$ if the LOS path of BS i is unobstructed, $b_k^{(i)}=0$ otherwise; $\mathbf{d}_\mathrm{tx}(\psi_k)\in\mathbb{C}^{M_\mathrm{tx}}$ is the BS array response vector with $\|\mathbf{d}_\mathrm{tx}(\psi_k)\|_2=1$; $\psi_k \triangleq \sin(\phi_k)= (x_k-L/2)/d_k$ is the spatial angle corresponding to the AoD (computed with respect to the perpendicular to the array) $\phi_k\in[-\Theta/2,\Theta/2]$ in slot k; the term $h_k\sim\mathcal{CN}(0,\sigma_h^2)$ is the complex channel gain of the LOS component, i.i.d. over slots, with $\sigma_h^2=1/\ell(d_k)$; $\ell(d_k)=[4\pi d_k]^2/\lambda_c^2$ denotes the distance-dependent path loss, as a function of the MU-BS distance $d_k=d(\phi_k)=D\sqrt{1+\tan(\phi_k)^2}$ (see Fig. 1); $\lambda_c=c/f_c$ is the wavelength at carrier frequency f_c.

Letting $G_\mathrm{tx}(\mathbf{c},\psi) = M_\mathrm{tx}|\mathbf{d}_\mathrm{tx}(\psi)^H\mathbf{c}|^2$ be the beamforming gain of the serving BS and $\Theta_\mathrm{tx} = \angle\mathbf{d}_\mathrm{tx}(\psi)^H\mathbf{c}$ be its phase, the signal received at the MU in slot k can be expressed as

$$\mathbf{y}_k = \sqrt{P_k}b_k^{(I_k)}h_k\sqrt{G_\mathrm{tx}(\mathbf{c}_k,\psi_k)}e^{j\Theta_\mathrm{tx}}\mathbf{x}_k + \mathbf{w}_k. \quad (5)$$

We use the sectored-antenna model, i.e., $G(\mathbf{c}^{(s)},\psi_k)/d(\phi_k)^2$ is constant within the main-lobe $\phi_k\in\Phi_s$, so that, letting $\Gamma \triangleq \frac{\lambda_c^2}{8\pi\sigma_w^2\Delta_s D}$, the average SNR when $\phi_k\in\Phi_s, b_k^{(I_k)}=1$ (alignment and no-blockage) can be shown to be

$$\mathrm{SNR}_k = \Gamma P_k, \quad (6)$$

This result is in line with the intuition that larger distances are achievable via smaller beamwidths, as also observed in [15]. If $\phi_k\notin\Phi_s$ or $b_k^{(I_k)}=0$ (mis-alignment or blockage), $\mathrm{SNR}_k = \rho\Gamma P_k$, where $\rho\in(0,1)$ is the side- to main-lobe gain ratio, which is numerically found from the gain pattern.

Finally, the blockage state $b_k^{(i)}$ is modeled as a Markov chain with transition probabilities

$$\mathbf{P}^{(i)}_{b\to b'} \triangleq \mathbb{P}(b_{k+1}^{(i)} = b'|b_k^{(i)} = b),\ \forall b,b'\in\{0,1\}. \quad (7)$$

The processes $\{b_k^{(i)}, k\geq 0\}, i\in\{1,2\}$ evolve independently of each other, with Markov dynamics (7). The independence assumption is motivated by the fact that the two BSs are on opposite sides of the road segment, hence they experience different types of obstructions between the MU and the BS. We now introduce the BT and DT operations.

BT phase: At the start of a BT phase, the BS selects a set of sectors $\hat{\mathcal{S}}_\mathrm{BT}$ over which it will send the beacons $\mathbf{x}_k$ for BT, and a target SNR, SNR_BT. The beacon transmission is done sequentially, using one slot for each sector in the set $\hat{\mathcal{S}}_\mathrm{BT}$. Therefore, the duration of the BT phase is $T_\mathrm{BT} \triangleq |\hat{\mathcal{S}}_\mathrm{BT}|+1$, which includes the last slot for feedback signaling from the MU to the BS. Let $i\in\{0,\dots,T_\mathrm{BT}-2\}$ be the ith timeslot during the BT phase, and $\hat{s}_i \in \hat{\mathcal{S}}_\mathrm{BT}$ be the sector covered

by the BS. At the MU, the received signal $\mathbf{y}_{k+i}$ is processed using a matched filter to generate the output

$$z_{\hat{s}_i} = \zeta(\mathbf{x}_{k+i}, \mathbf{y}_{k+i}) \triangleq \frac{|\mathbf{x}_{k+i}^H \mathbf{y}_{k+i}|^2}{N_0 W_{\rm tot} \|\mathbf{x}_{k+i}\|_2^2}. \quad (8)$$

Upon collecting the sequence $\{z_{\hat{s}}, \forall \hat{s} \in \hat{\mathcal{S}}_{\rm BT}\}$, the MU generates the feedback signal as

$$Y_k = \begin{cases} \hat{s}^* \triangleq \arg\max_{\hat{s}\in\hat{\mathcal{S}}_{\rm BT}} z_{\hat{s}}, & \max_{\hat{s}\in\hat{\mathcal{S}}_{\rm BT}} z_{\hat{s}} > \eta_{BT}, \\ \emptyset, & \max_{\hat{s}\in\hat{\mathcal{S}}_{\rm BT}} z_{\hat{s}} \le \eta_{BT}. \end{cases} \quad (9)$$

In other words, if all the matched filter outputs are below a threshold η_{BT}, the feedback $\emptyset$ is reported, indicating that no beam is deemed sufficient to carry data transmission, either due to blockage ($b_k^{(I)} = 0$), or mis-alignment ($Z_k \notin \hat{\mathcal{S}}_{\rm BT}$). Otherwise, the ID of the strongest beam $\hat{s}^*$ is reported.

DT phase: At the start of the DT phase, the BS selects a sector $\hat{s} \in \mathcal{S}$ over which it performs DT for $T_{\rm DT} - 1$ slots, along with a target average SNR at the receiver $\mathrm{SNR}_{\rm DT}$ and a target transmission rate $\bar{R}_{\rm DT}$; an additional slot is used for the feedback signal from the MU to the BS, as described below, so that the overall duration of the DT phase is $T_{\rm DT}$. We assume that a fixed fraction $\kappa \in (0,1)$ out of L symbols in each slot is used for channel estimation. Then, under alignment ($s = \hat{s}$ and $b_I = 1$), and assuming that channel estimation errors are negligible compared to the noise level (which can be achieved with a sufficiently long pilot sequence κL), from the signal model (5), we find that outage occurs if

$$W_{\rm tot} \log_2(1 + |h_k|^2 \ell(d_k) \mathrm{SNR}_{\rm DT}) < \bar{R}_{\rm DT}, \quad (10)$$

(note that $\mathbb{E}[|h_k|^2 \ell(d_k)] = 1$) yielding the outage probability

$$\mathbb{P}_{\rm OUT}(\bar{R}_{\rm DT}, \mathrm{SNR}_{\rm DT}) = 1 - \exp\Big\{ - \mathrm{SNR}_{\rm DT}^{-1} (2^{\frac{\bar{R}_{\rm DT}}{W_{\rm tot}}} - 1) \Big\}.$$

In this paper, we design $\bar{R}_{\rm DT}$ based on the notion of $\epsilon-$outage capacity, i.e., $\bar{R}_{\rm DT}$ is the largest rate such that $\mathbb{P}_{\rm OUT}(\bar{R}_{\rm DT}, \mathrm{SNR}_{\rm DT}) \le \epsilon$, for a target outage probability $\epsilon < 1$. Setting $\mathbb{P}_{\rm OUT} = \epsilon$, this can be expressed as

$$\bar{R}_{\rm DT} = C_\epsilon(\mathrm{SNR}_{\rm DT}) = W_{\rm tot} \log_2 \left(1 - \mathrm{SNR}_{\rm DT} \ln(1-\epsilon)\right), \quad (11)$$

so that the average throughput is

$$\mathcal{T}(\epsilon, \mathrm{SNR}_{\rm DT}) \triangleq (1-\kappa)(1-\epsilon) C_\epsilon(\mathrm{SNR}_{\rm DT}), \quad (12)$$

where $(1-\kappa)$ takes into account the overhead due to channel estimation. Subsequently, we select ϵ to maximize $\mathcal{T}$, i.e., given $\mathrm{SNR}_{\rm DT}$, ϵ is chosen as the unique fixed point of $\mathrm{d}\mathcal{T}(\epsilon, \mathrm{SNR}_{\rm DT})/\mathrm{d}\epsilon = 0$. We denote the corresponding throughput maximized over ϵ as $\mathcal{T}^*(\mathrm{SNR}_{\rm DT})$.

We envision a mechanism in which the pilot signal transmitted in the second last slot of the DT phase (the most recent) is used to generate the binary feedback signal $Y \in \{\hat{s}, \emptyset\}$, transmitted by the MU back to the BS in the last slot of the DT phase. Similarly to the BT feedback,

$$Y_k = \begin{cases} \hat{s}, & \zeta(\mathbf{x}^{(p)}_{k+T_{\rm DT}-2}, \mathbf{y}^{(p)}_{k+T_{\rm DT}-2}) > \eta_{DT} \\ \emptyset, & \zeta(\mathbf{x}^{(p)}_{k+T_{\rm DT}-2}, \mathbf{y}^{(p)}_{k+T_{\rm DT}-2}) \le \eta_{DT}, \end{cases} \quad (13)$$

based on the pilot signal $\mathbf{x}^{(p)}_{k+T_{\rm DT}-2}$ (of duration κL) and on the corresponding signal $\mathbf{y}^{(p)}_{k+T_{\rm DT}-2}$ received on the second last slot of the DT phase, so that $Y = \hat{s}$ denotes beam-alignment, whereas $Y = \emptyset$ denotes loss of alignment due to either mobility of the MU or blockage. For both BT and DT, the feedback distribution is computed in closed-form in [1].

III. POMDP Formulation

We now formulate the problem of jointly optimizing the BT, DT and HO strategy as a POMDP, defined next.

States: the state is denoted as $u_k \triangleq (Z_k, I_k, b_k^{(1)}, b_k^{(2)}) \in \mathcal{U}$ taking values from the set $\mathcal{U} = (\mathcal{S} \times \{1,2\} \times \{0,1\}^2)$, where $Z_k \in \mathcal{S}$ is the sector occupied by the MU, $I_k \in \{1,2\}$ is the index of the serving BS, and $b_k^{(i)} \in \{0,1\}$ for $i \in \{1,2\}$ is the blockage state. We add the absorbing state $\bar{s}$ to denote the fact that the MU exited the coverage area of the two BSs, so that the overall state space is $\bar{\mathcal{U}} = \mathcal{U} \cup \{\bar{s}\}$.

Actions: the serving BS can perform either BT, DT or HO actions. However, differently from standard POMDPs in which each action takes one slot, in this paper we generalize the model to actions taking multiple slots, as explained next.

Under action HO, the other BS becomes the serving one for the successive time-slots, until HO is chosen again. Its duration is denoted as $T_{\rm HO}$, modeling the delay to coordinate the transfer of the data traffic between the two BSs.

Under action BT, the serving BS chooses the set $\hat{\mathcal{S}}_{\rm BT}$ of sectors to scan and the target SNR $\mathrm{SNR}_{\rm BT}$. The duration of the BT action is $T_{\rm BT} = |\hat{\mathcal{S}}_{\rm BT}| + 1$: $|\hat{\mathcal{S}}_{\rm BT}|$ slots for scanning the set of sectors $\hat{\mathcal{S}}$, and one slot for the feedback from the MU to the serving BS.

Under action DT, the serving BS selects the sector $\hat{s}$ covered, the duration $T_{\rm DT} \ge 2$, and the target SNR $\mathrm{SNR}_{\rm DT}$ of the data communication session. The transmission power is then determined via (6), and the transmission rate is given by (11) to achieve ϵ-outage capacity, so that the resulting expected throughput (in case of LOS and correct alignment) is $\mathcal{T}^*(\mathrm{SNR}_{\rm DT})$. The duration of the data communication session $T_{\rm DT}$ includes the second last slot to generate the feedback signal, which is fed back to the BS in the last slot.

We denote the action as the 4-tuple $a = (c, \hat{\mathcal{S}}_c, \mathrm{SNR}_c, T_c)$, where $c \in \{\mathrm{HO}, \mathrm{BT}, \mathrm{DT}\}$ is the action class. For HO, we set $\hat{\mathcal{S}}_{\rm HO} = \emptyset$ and $\mathrm{SNR}_{\rm HO} = 0$. We denote the action space as $\mathcal{A}$.

Observations: upon selecting action $A_k \in \mathcal{A}$ of duration T in slot k and executing it in state $u_k \in \mathcal{U}$, the BS observes Y_k from the set $\mathcal{Y} = \mathcal{S} \cup \{\emptyset\} \cup \{\bar{s}\}$. The observation signal $Y_k = \bar{s}$ denotes that the MU exited the coverage area of the two BSs, hence the episode terminates; otherwise, Y_k denotes the feedback signal after the action is completed, as described earlier for the BT and DT actions in (9) and (13) (we set $Y_k = \emptyset$ under the HO action).

Transition, Observation probabilities: Let $\mathbb{P}(u', y | u, a) \triangleq \mathbb{P}(U_{k+T} = u', Y_k = y | U_k = u, A_k = a)$ be the probability of moving from state $u \in \mathcal{U}$ to state $u' \in \bar{\mathcal{U}}$ and observing $y \in \mathcal{Y}$ under action $a \in \mathcal{A}$ of duration T. Note that these probabilities are a function of the duration T of the selected action a, and can be computed in closed-form based on the feedback distribution and state transition probabilities (see [1]).

Costs and Rewards: for every state action pair (u, a), we let $r(u,a)$ and $e(u,a)$ be the expected number of bits transmitted from the BS to the MU and the expected energy cost, respectively. Under the HO and BT actions, we have that $r(u,a) = 0$ (since no bits are transmitted under these actions). On the other hand, under the DT action $a = (\mathrm{DT}, \{\hat{s}\}, \mathrm{SNR}, T_{\rm DT})$ (of duration $T_{\rm DT}$, SNR SNR, over sector $\hat{s}$), the expected throughput

in the tth communication slot is $\mathcal{T}^*(\mathrm{SNR})$, provided that there is correct alignment and no blockage ($Z_{k+t}{=}\hat{s}$ and $b^{(I)}_{k+t}{=}1$); otherwise, outage occurs and the expected throughput is zero. Hence, the total expected traffic delivered over the entire communication session is

$$r((s,I,b_1,b_2),(\mathrm{DT},\{\hat{s}\},\mathrm{SNR},T_{\mathrm{DT}})) \quad (14)$$
$$=\mathcal{T}^*(\mathrm{SNR})\sum_{t=0}^{T_{\mathrm{DT}}-2}\mathbb{P}(Z_{k+t}{=}\hat{s},b^{(I)}_{k+t}{=}1|Z_k{=}s,b^{(I)}_k{=}b_I).$$

The energy cost under action a with SNR SNR is expressed from (6) as (note that $\mathrm{SNR}{=}0$ and $e(u,a){=}0$ under HO)

$$e(u,a)=\frac{\Delta_{\mathrm{t}}}{\Gamma}\mathrm{SNR}(T-1). \quad (15)$$

Note that the last slot is reserved to the feedback transmission, which incurs no energy cost for the BS. We opt for a Lagrangian formulation to trade-off cost $e(u,a)$ and reward $r(u,a)$, and we define $\mathcal{L}(u,a){=}r(u,a){-}\lambda e(u,a)$ for $\lambda{\geq}0$.

Policy and Belief updates: Since the agent cannot directly observe the system state u_k, we introduce the notion of *belief* $\beta\in\mathcal{B}$, i.e., the probability distribution over system states, given the information collected so far at the BS. Given β, the serving BS selects an action a according to a policy $a=\pi(\beta)$, part of our design in Sec. IV; then, upon executing the action a and receiving the feedback signal y, the BS updates the belief for the next decision interval according to Bayes' rule as

$$\beta'(u')=\mathbb{P}(u'\mid y,a,\beta)=\frac{\sum_{u\in\mathcal{U}}\mathbb{P}(u',y|u,a)\beta(u)}{\sum_{u\in\mathcal{U}}\sum_{u''\in\bar{\mathcal{U}}}\mathbb{P}(u'',y|u,a)\beta(u)},$$

where $\mathbb{P}(u',y|u,a)$ is the conditional joint state transition and observation probability [1].

IV. Optimization Problem

Our goal is to determine a policy π (i.e., a map from beliefs to actions) that maximizes a trade-off between throughput and average power, $\bar{V}^{\pi}{\triangleq}\bar{T}^{\pi}{-}\lambda\bar{P}^{\pi}$, starting from a given initial belief $\beta_0=\beta_0^*$ at time 0. Using Little's Theorem [16], these metrics can be expressed as

$$\bar{T}^{\pi}\triangleq\frac{\bar{R}^{\pi}_{\mathrm{tot}}(\beta_0^*)}{\bar{D}_{\mathrm{tot}}(\beta_0^*)},\ \bar{P}^{\pi}\triangleq\frac{\bar{E}^{\pi}_{\mathrm{tot}}(\beta_0^*)}{\bar{D}_{\mathrm{tot}}(\beta_0^*)},\ \bar{V}^{\pi}\triangleq\frac{\bar{V}^{\pi}_{\mathrm{tot}}(\beta_0^*)}{\bar{D}_{\mathrm{tot}}(\beta_0^*)}, \quad (16)$$

where $\bar{D}_{\mathrm{tot}}(\beta_0^*)$ is the expected episode duration, function of the mobility process but independent of policy π,

$$[\bar{R}^{\pi}_{\mathrm{tot}}(\beta),\bar{E}^{\pi}_{\mathrm{tot}}(\beta)]\triangleq\mathbb{E}_{\pi}\Big[\sum_{t=0}^{\infty}[r(u_t,a_t),e(u_t,a_t)]\Big|\beta_0{=}\beta\Big]$$

are the total expected number of bits transmitted and the total expected energy cost during an episode,[1] and

$$\bar{V}^{\pi}_{\mathrm{tot}}(\beta){=}\bar{R}^{\pi}_{\mathrm{tot}}(\beta){-}\lambda\bar{E}^{\pi}_{\mathrm{tot}}(\beta)=\mathbb{E}_{\pi}\Big[\sum_{t=0}^{\infty}\mathcal{L}(u_t,a_t)\Big|\beta_0{=}\beta\Big].$$

Then, the optimization problem starting from the initial belief $\beta_0=\beta_0^*$ is expressed as

$$\max_{\pi}\ \bar{V}^{\pi}(\beta_0^*){=}\frac{1}{\bar{D}_{\mathrm{tot}}(\beta_0^*)}\max_{\pi}\bar{V}^{\pi}_{\mathrm{tot}}(\beta_0^*).$$

[1]Note that the convergence of these series is guaranteed by the presence of the absorbing state $\bar{s}$, i.e., the MU exits the coverage area at some point.

It is well known that the optimal value function uniquely satisfies Bellman's optimality equation [2] $V^*{=}H[V^*]$, where we have defined the operator $\hat{V}{=}H[V]$ as

$$\hat{V}(\beta){=}\max_{a\in\mathcal{A}}\sum_{u\in\mathcal{U}}\beta(u)\Big[\mathcal{L}(u,a){+}\sum_{y,u'}\mathbb{P}(u',y|u,a)V(\mathbb{B}(y,a,\beta))\Big],$$

$\forall\beta\in\mathcal{B}$, and the maximizer is the optimal policy $\pi^*(\beta)$. The optimal value function V^* can be arbitrarily well approximated via the value iteration algorithm $V_{n+1}{=}H[V_n]$, where $V_0(\beta){=}0,\forall\beta$. Moreover, V_n is a piece-wise linear and concave function [2], so that it can be expressed by a finite set of hyperplanes $\mathcal{Q}_n\equiv\{\alpha_{n,i}\}_{i=1}^{A_n}$ of cardinality A_n, such that

$$V_n(\beta)=\max_{\alpha\in\mathcal{Q}_n}\beta\cdot\alpha, \quad (17)$$

where $\beta{\cdot}\alpha{=}\sum_u\beta(u)\alpha(u)$ denotes inner product. Each hyperplane $\alpha{\in}\mathcal{Q}_n$ is associated with an action $a_\alpha{\in}\mathcal{A}$, so that the maximizing hyperplane α^* in (17) defines the policy $\pi_n(\beta){=}a_{\alpha^*}$. It has been shown that $\mathcal{Q}_n$ grows doubly exponentially with the number of iterations, $A_{n+1}{=}|\mathcal{Q}_{n+1}|{=}|\mathcal{A}|^{|\mathcal{Y}|^n}$ [17]. For this reason, computing optimal policies for POMDPs is an intractable problem for any reasonably sized task. This calls for approximate solution techniques, e.g., PERSEUS [2], which we introduce next.

A. Point-based Value Iteration (PBVI) for POMDPs

PERSEUS [2] is an approximate PBVI algorithm for POMDPs. The key idea is to define an approximate backup operator $\tilde{H}[\cdot]$ (in place of $H[\cdot]$), restricted to a discrete subset of belief points in $\tilde{\mathcal{B}}$, chosen as representative of the entire belief space $\mathcal{B}$; in other words, for a given value function $\tilde{V}_n$ at stage n, PERSEUS builds a value function $\tilde{V}_{n+1}=\tilde{H}[\tilde{V}_n]$ that improves the value of all belief points $\beta\in\tilde{\mathcal{B}}$, without regard for the belief points outside of this discrete set, $\beta\notin\tilde{\mathcal{B}}$. The goal of the algorithm is to provide a $|\tilde{\mathcal{B}}|$-dimensional set of hyperplanes $\alpha\in\mathcal{Q}$ and associated actions a_α. Given such set, the value function at any other belief point $\beta\in\mathcal{B}$ is then approximated via (17) as $\tilde{V}(\beta)=\beta\cdot\alpha^*$, where $\alpha^*=\arg\max_{\alpha\in\mathcal{Q}}\beta\cdot\alpha$, which defines an approximately optimal policy $\pi(\beta)=a_{\alpha^*}$.

The approximate backup operation of PERSEUS is given by Algorithm 1, which takes as input a set of hyperplanes $\mathcal{Q}_n$ and the corresponding actions, and outputs a new set $\mathcal{Q}_{n+1}$ along with their corresponding actions. To do so: in line 3, a belief point is chosen randomly from $\tilde{\mathcal{B}}_{\mathrm{temp}}$; in lines 4-5, the hyperplane associated with each action $a\in\mathcal{A}$ is computed; in particular, line 4 computes the hyperplane associated with the future value function $V_n(\mathbb{B}(y,a,\beta))$, for each possible observation y resulting in the belief update $\mathbb{B}(y,a,\beta)$; line 5 instead performs the backup operation to determine the new one-step lookahead hyperplane associated with each action; line 6 determines the optimal action that maximizes the value function for the current belief, yielding overall the value iteration update $V_{n+1}(\beta)=\max_a\mathbb{E}_{U,Y|a,\beta}[\mathcal{L}(U,a)+V_n(\mathbb{B}(Y,a,\beta))]$; in lines 7-10, the new hyperplane and the associated action is added to the set $\mathcal{Q}_{n+1}$, but only if it yields an improvement in the value function $V_{n+1}(\beta)>\tilde{V}_n(\beta)$; otherwise, the previous hyperplane is used; finally, lines 11-12 update the set of unimproved beliefs based on the newly added hyperplane; only the belief points that have not been improved are part of the next iterations of the algorithm. Overall, the algorithm

Algorithm 1: function PERSEUS

input : $\tilde{\mathcal{B}}$, $\mathcal{Q}_n$, $\{a_\alpha^n, \alpha \in \mathcal{Q}_n\}$

Init: $\tilde{V}_{n+1}(\tilde{\beta}) = -\infty, \forall \tilde{\beta} \in \tilde{\mathcal{B}}$; $\tilde{\mathcal{B}}_{\text{temp}} \equiv \tilde{\mathcal{B}}$, $\mathcal{Q}_{n+1} = \emptyset$;
$\tilde{V}_n(\tilde{\beta}) \leftarrow \max_{\alpha \in \mathcal{Q}_n} \tilde{\beta} \cdot \alpha$, maximizer $\alpha_{\tilde{\beta}}$, $\forall \tilde{\beta} \in \tilde{\mathcal{B}}$;

while $\tilde{\mathcal{B}}_{\text{temp}} \neq \emptyset$ **do** // Unimproved beliefs

  Sample β from $\tilde{\mathcal{B}}_{\text{temp}}$; For each action a, solve

  $\alpha_{y,a}^* = \arg\max_{\alpha \in \mathcal{Q}_n} \mathbb{B}(y, a, \beta) \cdot \alpha$, $\forall y \in \mathcal{Y}$ and

  $\hat{\alpha}_a^*(u) = \mathcal{L}(u, a) + \sum_{\hat{u}, y} \mathbb{P}(\hat{u}, y | u, a) \alpha_{y,a}^*(\hat{u})$, $\forall u$;

  Solve $V_{n+1}(\beta) = \max_{a \in \mathcal{A}} \beta \cdot \hat{\alpha}_a^*$ and maximizing action a^* and hyperplane $\hat{\alpha} = \hat{\alpha}_{a^*}^*$;

  if $V_{n+1}(\beta) > \tilde{V}_n(\beta)$ **then** // $\hat{\alpha}$ improves value

    $\mathcal{Q}_{n+1} \leftarrow \mathcal{Q}_{n+1} \cup \{\hat{\alpha}\}$; $a_{\hat{\alpha}}^{n+1} = a^*$ // add $\hat{\alpha}$ to $\mathcal{Q}_{n+1}$ and define action associated to $\hat{\alpha}$;

  else // keep previous hyperplane α_β

    $\hat{\alpha} = \alpha_\beta$; $\mathcal{Q}_{n+1} \leftarrow \mathcal{Q}_{n+1} \cup \{\alpha_\beta\}$; $a_{\alpha_\beta}^{n+1} = a_{\alpha_\beta}^n$;

  $\tilde{V}_{n+1}(\tilde{\beta}) \leftarrow \max\{\tilde{\beta} \cdot \hat{\alpha}, \tilde{V}_{n+1}(\tilde{\beta})\}, \forall \tilde{\beta} \in \tilde{\mathcal{B}}$;
  // unimproved beliefs

  $\tilde{\mathcal{B}}_{\text{temp}} \leftarrow \{\tilde{\beta} \in \tilde{\mathcal{B}}_{\text{temp}} : \tilde{V}_{n+1}(\tilde{\beta}) < \tilde{V}_n(\tilde{\beta})\}$;

return $\mathcal{Q}_{n+1}$, $\{a_\alpha^{n+1}, \forall \alpha \in \mathcal{Q}_{n+1}\}$ // new hyperplanes and associated actions

guarantees monotonic improvements of the value function in the set $\tilde{\mathcal{B}}$, and continues until all beliefs have been improved and $\tilde{\mathcal{B}}_{\text{temp}}$ is empty. Algorithm 1 is then executed iteratively, until convergence of the value function to a fixed point.

To generate $\tilde{\mathcal{B}}$, we employ the *Stochastic simulation and exploratory action* (SSEA) algorithm [17]. After initializing $\mathcal{B}_0$, at iteration n, SSEA iteratively performs a one step forward simulation with each action in the action set, thus producing new beliefs $\{\beta_a, \forall a \in \mathcal{A}\}$; hence, it computes the L1 distance between each new belief point β_a and its closest neighbor in $\mathcal{B}_n$, and adds the belief point β_{a^*} farthest away from $\mathcal{B}_n$, so as to provide a wider coverage of the belief space. This expansion is performed multiple times to obtain $\tilde{\mathcal{B}}$.

After returning the set of hyperplanes $\mathcal{Q}_{n+1}$ and the associated actions $\{a_\alpha^{n+1}, \forall \alpha \in \mathcal{Q}_{n+1}\}$, the (approximately) optimal action when operating under the belief β can be computed as

$$\pi^*(\beta) = a_{\alpha^*}^{n+1}, \text{ where } \alpha^* = \arg \max_{\alpha \in \mathcal{Q}_{n+1}} \beta \cdot \alpha.$$

In Fig. 2, we plot a time-series of the evolution of state variables for a portion of an episode executed under the PERSEUS-based policy (Algorithm 1). The parameters used are listed in Table 1. Initially, the MU is known to be in sector Z_0=1, with LOS conditions for both BSs ($b_0^{(1)}$=$b_0^{(2)}$=1). We show a time-series for the sector index Z_k, index of the serving BS I_k, its blockage state $b_k^{(I_k)}$, the action class $c \in \{\text{DT}, \text{BT}, \text{HO}\}$, the BT feedback Y_{BT} as defined in (9), and the DT feedback Y_{DT} as defined in (13). The action space for the DT time is set as $T_{\text{DT}} \in \{10, 20, 40\}$ and the power levels are set as $P_{\text{BT}}, P_{\text{DT}} \in \{0, 10, 20, 30, 40\}$ (dBm). It can be observed in the figure that at 0.238s, 0.246s and 0.287s, NACKs are received after executing the DT action. After each one of these NACKs, the policy executes the BT action. If the BT feedback $Y_{\text{BT}} \neq \emptyset$, then DT is performed; otherwise, blockage is detected and the HO action is executed. Next, we will present a heuristic policy that mimics this behavior.

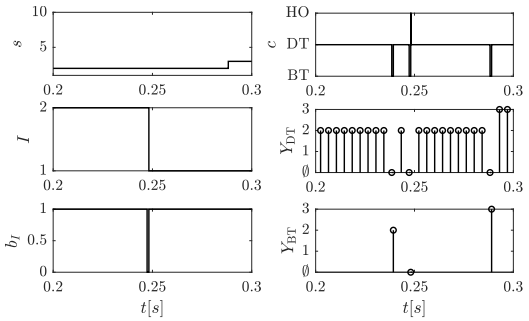

Fig. 2: Execution of policy π^*.

B. Heuristic Policy

Note that Algorithm 1 incurs a huge computational cost especially for POMDP with large state and action spaces (hence large number of representative belief points). To remedy this, we propose a finite state machine based heuristic policy (FSM-HEU) that will be shown numerically to achieve *near-optimal* performance. The key idea of FSM-HEU is that it selects the BT/DT/HO actions based solely on the last action executed and its observation signal, but not on the belief β_k. The behavior of this scheme can thus be described as a finite-state machine, depicted in Fig. 3 and described next.

If the last action executed was a BT action, and the feedback signal is $Y = \hat{s}$ (see (9)), then the BS detects the strongest beam $\hat{s}$; hence the next action selected is DT over sector $\hat{s}$ (the strongest detected), of fixed duration T_{DT}. On the other hand, if the feedback signal is $Y = \emptyset$, the BS detects blockage and performs handover to the non-serving BS (action HO).

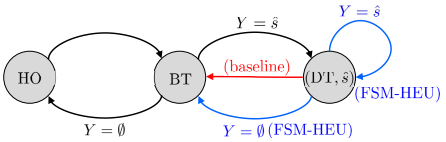

Fig. 3: Finite state machine based on the observation signal Y. Black lines represent transitions under both FSM-HEU and baseline; blue and red lines represent transitions under FSM-HEU and baseline only, respectively.

If the last action executed was DT on sector $\hat{s}$, and the feedback signal is ACK (Y=$\hat{s}$, see (13)), then the BS infers that the signal is still sufficiently strong to continue DT on the same sector, and the same action is selected; otherwise (NACK received, Y=$\emptyset$), the BS detects a loss of alignment, hence the BT action with exhaustive search is selected.

Finally, if the last action executed was HO, then the new serving BS executes BT via exhaustive search to locate the MU. This procedure continues until the episode terminates.

To study its performance, note that the underlying system state U_k and action A_k form a Markov chain. Letting $\mathbb{P}(a'|y, a)$ be the probability of generating the new action a', given previous action a and observation y, as given by the finite-state machine of Fig. 3, the value fuction $V(u, a), \forall u, a$ is obtained by solving the following system of linear equations

$$V(u, a) = \mathcal{L}(u, a) + \sum_{y, u', a'} \mathbb{P}(u', y | u, a) \mathbb{P}(a' | y, a) V(u', a'), \forall u, a.$$

V. Numerical Results

In this section, we perform a numerical evaluation of the various algorithms proposed in this paper, with simulation parameters listed in Table 1.The blockage transition probabilities given in the table correspond to steady state blockage

Parameter	Symbol	Value
Number of BS antennas	M_{tx}	128
Angular BS coverage	Θ	90°
Slot duration	Δ_t	100μs
Distance of road to BS	D	20m
Bandwidth	W_{tot}	100MHz
Carrier frequency	f_c	30GHz
Noise psd	N_0	$-$163dBm/Hz
Fraction of DT slot for channel-estimation/hypothesis-testing	κ	0.01
HO delay	T_{HO}	1 slot
LOS to blockage transition prob.	$\mathbf{P}_{1\to0}$	1.25×10^{-4}
Blockage to LOS transition prob.	$\mathbf{P}_{0\to1}$	5×10^{-4}
MU average speed	μ_v	30m/s
MU speed standard deviation	σ_v	10
MU mobility memory parameter	γ	0.2

TABLE 1: Simulation parameters.

probabilities $\pi_B^{(1)}=\pi_B^{(2)}=0.2$ and average blockage duration of 0.2ms. Using the throughput metric defined in (16), the average spectral efficiency is computed as $\bar{T}^{\pi}/W_{\text{tot}}$[bps/Hz]. We compare the performance of the proposed policies to a baseline scheme which performs periodic BT, unless blockage is detected (in which case it executes HO, see Fig. 3).

In Fig. 4, we depict the average spectral efficiency against the average power consumption. For the FSM-HEU and baseline policies, we set $T_{\text{DT}}=10$, and $P_{\text{BT}}=P_{\text{DT}}$ is varied from 0dBm to 40dBm. The upper-bound shown in the figure is obtained by a genie-aided policy that always executes DT with perfect knowledge of the state (s, I, b_1, b_2) and hence its throughput performance can be upper bounded by $(1-\pi_B^{(1)}\pi_B^{(2)})\mathcal{T}^*(\text{SNR}_{\text{DT}})$, i.e., it is $\mathcal{T}^*(\text{SNR}_{\text{DT}})$ unless there is no LOS under both BSs (with steady-state probability $\pi_B^{(1)}\pi_B^{(2)}$) whereas its power consumption is given as $(1-\pi_B^{(1)}\pi_B^{(2)})P_{\text{DT}}$. Note that this upper-bound is not attainable since it is found by assuming perfect knowledge of the state and ignoring the inefficiencies due to the time required to perform handover and transmit feedback. The PERSEUS-based policy π^* yields the best performance with negligible performance gap with respect to the upper-bound. It shows a performance gain of up to 11% and 55% compared to FSM-HEU and baseline, respectively. However, the baseline policy yields up to 50% degraded performance compared to FSM-HEU: in fact, the baseline scheme neglects the DT feedback and instead performs periodic BT, thus incurring significant overhead. We observe that the curves corresponding to analysis and the one based on simulation (based on the Gauss-Markov mobility model and beam design via [13]) closely match, thereby showing that the model introduced in the paper provides good abstraction of more realistic settings.

VI. Conclusions

In this paper, we have investigated the design of beam-training/data-transmission/handover strategies for mm-wave vehicular networks. The mobility and blockage dynamics have been leveraged to obtain the approximately optimal policy via a POMDP formulation and its solution via a point-based value iteration (PBVI) algorithm based on PERSEUS [2]. Inspired by it, we have proposed a heuristic policy, which provides low computational alternatives to PBVI and exhibits performance comparable to the optimal policy obtained via PBVI. Our numerical results demonstrate the importance of an adaptive design to tackle the highly dynamic environments caused by mobility and blockages in vehicular networks.

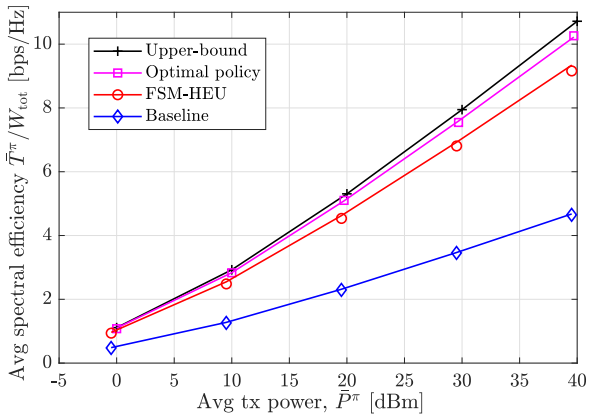

Fig. 4: Average spectral efficiency versus average power consumption: analytical curves based on the sectored antenna and mobility model (continuous lines) and simulation using analog beamforming and Gauss-Markov mobility (markers).

References

[1] M. Hussain, M. Scalabrin, M. Rossi, and N. Michelusi, "Mobility and Blockage-aware Communications in Millimeter-Wave Vehicular Networks," 2020, submitted to IEEE Transactions on Vehicular Technology. [Online]. Available: http://arxiv.org/abs/2002.11210

[2] M. T. J. Spaan and N. Vlassis, "Perseus: Randomized point-based value iteration for pomdps," *J. Artif. Int. Res.*, vol. 24, no. 1, pp. 195–220, Aug. 2005.

[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, 2016.

[4] N. Michelusi and M. Hussain, "Optimal beam-sweeping and communication in mobile millimeter-wave networks," in *2018 IEEE International Conference on Communications (ICC)*, May 2018, pp. 1–6.

[5] Z. Marzi, D. Ramasamy, and U. Madhow, "Compressive channel estimation and tracking for large arrays in mm-wave picocells," *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, no. 3, pp. 514–527, April 2016.

[6] V. Va, J. Choi, T. Shimizu, G. Bansal, and R. W. Heath, "Inverse multipath fingerprinting for millimeter wave v2i beam alignment," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 5, pp. 4042–4058, May 2018.

[7] M. Hussain and N. Michelusi, "Energy-efficient interactive beam alignment for millimeter-wave networks," *IEEE Transactions on Wireless Communications*, vol. 18, no. 2, pp. 838–851, Feb 2019.

[8] V. Va, T. Shimizu, G. Bansal, and R. W. Heath, "Beam design for beam switching based millimeter wave vehicle-to-infrastructure communications," in *2016 IEEE ICC*, 2016, pp. 1–6.

[9] M. Scalabrin, N. Michelusi, and M. Rossi, "Beam training and data transmission optimization in millimeter-wave vehicular networks," in *2018 IEEE Globecom*, Dec 2018, pp. 1–7.

[10] V. Va, T. Shimizu, G. Bansal, and R. W. Heath, "Online learning for position-aided millimeter wave beam training," *IEEE Access*, vol. 7, pp. 30 507–30 526, 2019.

[11] M. Hussain and N. Michelusi, "Second-best beam-alignment via bayesian multi-armed bandits," in *2019 IEEE Globecom*, 2019, to appear. [Online]. Available: http://arxiv.org/abs/1906.04782

[12] A. Alkhateeb, I. Beltagy, and S. Alex, "Machine learning for reliable mmwave systems: Blockage prediction and proactive handoff," in *IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, Nov 2018, pp. 1055–1059.

[13] S. Noh, M. D. Zoltowski, and D. J. Love, "Multi-Resolution Codebook and Adaptive Beamforming Sequence Design for Millimeter Wave Beam Alignment," *IEEE Transactions on Wireless Communications*, vol. 16, no. 9, pp. 5689–5701, Sep. 2017.

[14] T. Bai and R. W. Heath, "Coverage analysis for millimeter wave cellular networks with blockage effects," in *2013 IEEE Global Conference on Signal and Information Processing*, Dec 2013, pp. 727–730.

[15] M. Giordani, M. Mezzavilla, and M. Zorzi, "Initial access in 5g mmwave cellular networks," *IEEE Communications Magazine*, vol. 54, no. 11, pp. 40–47, November 2016.

[16] J. D. C. Little and S. Graves, *Little's Law*, 07 2008, pp. 81–100.

[17] J. Pineau, G. Gordon, and S. Thrun, "Anytime point-based approximations for large pomdps," *J. Artif. Int. Res.*, vol. 27, no. 1, pp. 335–380, Nov. 2006.