

On Velocity-based Association Policies for Multi-tier 5G Wireless Networks

Pierre Popineau¹, Sanket S. Kalamkar², and François Baccelli¹

Affiliation: ¹INRIA-ENS, Paris, France, ²Qualcomm Inc., San Diego, USA.

Abstract—Mobility is a key challenge for beam management in 5G cellular networks due to the overhead incurred at beam switching and base station (BS) handover events. This paper focuses on a network that has a multi-tier structure with two types of BSs operating in the same frequency bands, namely macro BSs that are sparser but with higher transmit power, and micro BSs that are denser and with lower transmit power. We propose a downlink user association policy which is a function of the user mobility. Typically, high mobility users should associate with macro BSs so as to incur less beam switching overhead, whereas low mobility ones should be associated with micro BSs. The main contribution of the paper is a formalization of the optimal threshold association policy, when the optimality is understood with respect to the effective Shannon rate. The analysis is based on stochastic geometry and on an exact representation of the effective Shannon rate of the typical user in this beamforming multi-tier context. Two models are discussed. The simplest one focuses on a single-user optimization problem. We also discuss a more realistic model with bandwidth sharing between all users in the cell. Finally, we identify the mobility and user-density patterns where the velocity-based threshold association policy outperforms the classical best mean power association policy.

I. INTRODUCTION

A key feature of 5G is beam-based communications, which allows one to concentrate the transmit energy in a specific direction and improve the signal power and simultaneously reduce the interference [1]. Although more crucial for millimeter wave frequencies, a beam-centric design also helps improve network performance at sub-6 GHz frequencies. However, the key components and procedures of beam management, namely beam selection, beam failure detection, and beam switching, involve different signaling overheads which may degrade the network performance, especially for highly mobile users [2]. For instance, the signal-to-interference-plus-noise ratio (SINR) could be improved by using narrower beams, but this benefit brings a significant cost in terms of more frequent beam switching and beam misalignment.

The setting of the present paper is that of a network with a multi-tier architecture, namely featuring different types of base stations (BSs) using the same frequency bands, where differences are only in terms of spatial density and transmit powers. The archetype of such an architecture is that with micro cells overlaying macro cells. It is well known that this can quite significantly increase the network performance [3]. When such a multi-tier network is accessed by users with different mobility patterns, e.g., pedestrians, vehicles, and high-speed trains, an important question is that of the

association policy to be used. The answer to this question is fundamentally impacted by the beam management used in each tier. For instance, although connecting to micro cells may offer a better coverage and throughput to a mobile user due to some dense deployment of BSs, if this user is moving at a sufficiently high speed, frequent BS handovers and beam switching events combined with a higher chance of beam misalignment may significantly degrade the effective rate obtained by the user.

The present paper addresses this question and describes a mobility-aware association policy within this 5G multi-tier setting. It identifies the optimal association policy within a class of threshold policies which are particularly simple to implement. Optimality is in a sense to be defined precisely in the paper; it is based on a system-level view of different tradeoffs associated with beam management, mobility, load on a cell, spectral efficiency, and the multi-tier parameters. This system-level view allows one to define the notion of the typical user. The optimization coincides with the maximization of the expected effective rate that this typical user gets.

A. Related work

Stochastic geometry is widely used to study several types of wireless networks [4], [5], including multi-tier cellular networks [3]. There are several works that study the effect of user mobility on the optimal tier selection in cellular networks [6]–[8]. These works have provided important insights into how user mobility affects the performance metrics such as coverage probability, Shannon rate, and BS handover rate. While the works in [6], [7] focus on a single mobile user scenario, [8] considers the case of multiple mobile users who share the network resources. However, to the best of our knowledge, there is no work available on the beam management in a multi-tier network consisting of mobile users with different speed profiles. There exists one work [9] that has developed a mathematical optimization framework for the beam management in a 5G network and focused on the tradeoffs associated with mobility and beam management. But this last work focuses only on a single-user single-tier cellular network. Overall, the question of optimal user association policy in the presence of intra-cell mobility leading to beam switching, BS handovers, and resource sharing among multiple users is still unanswered.

II. SYSTEM MODEL

A. Network setup

Consider a two-tier downlink cellular network, where we call the first tier the *macro* tier and the second one the

E-mail: pierre.popineau@inria.fr, francois.baccelli@inria.fr.
This work was done when Sanket S. Kalamkar was at INRIA, Paris.

micro tier. We assume that BSs are located according to two independent homogeneous Poisson point processes (PPPs): the macro BSs form a PPP $\Phi_M \subset \mathbb{R}^2$ of intensity λ_M and the micro BSs a PPP Φ_μ of intensity λ_μ , with $\lambda_\mu > \lambda_M$. Mobile users (MUs) are located according to a stationary PPP Φ_u of intensity λ_u . Each MU moves on a straight line with an orientation chosen at random and independent of other MUs. Also, each MU travels at a constant velocity v , sampled from a given distribution f , independently for each user. Without loss of generality, using the homogeneity and isotropy of the PPP, we can assume that the typical MU is located at the origin and that its motion is along the x -axis.

We assume that all BSs are always active. Some of our results can be extended to the case where BSs with no users are shut off, but this does not fall in the scope of this work. The formulas of this section are presented for the macro tier, but they hold for the micro tier by swapping the roles of the two tiers.

B. Beamforming

We assume that macro BSs (resp. micro BSs) use directional beamforming to communicate with the typical MU. Each macro BS (resp. micro BS) has n_M (resp. n_μ) beams. Each beam has the same angular width $\psi_M = \frac{2\pi}{n_M}$ in the macro tier and $\psi_\mu = \frac{2\pi}{n_\mu}$ in the micro tier. The main lobe is restricted to the beamwidth. The antenna gain in the macro tier is assumed to be:

$$G^M(\theta) = \begin{cases} G_m^M & \text{if } |\theta| \leq \psi_M/2 \\ G_s^M & \text{else,} \end{cases}$$

where G_m^M is the main lobe gain and G_s^M is the side lobe gain. The probability that an MU lies within the main lobe of an interfering macro BS is $p_{M,m} = \frac{\psi_M}{2\pi} = \frac{1}{n_M}$, and that of an interfering micro BS is $p_{\mu,m} = \frac{1}{n_\mu}$. Thus the BS antenna gain $g_{M,x}$, w.r.t. the typical MU, of an interfering BS located at x in the macro tier is given by

$$g_{M,x} = \begin{cases} G_m^M & \text{w.p. } p_{M,m} \\ G_s^M & \text{w.p. } 1 - p_{M,m}. \end{cases} \quad (1)$$

C. Mobility-based beam misalignment

Due to intra-cell mobility, the MU has to reselect the beam when it moves from the coverage of one beam to another. Such a beam reselection in a 5G network occurs during a synchronization signal block (SSB) burst with period τ . If the MU moves out of the main lobe of its original connection beam (also called the reference beam) between two consecutive SSB bursts without selecting a new beam, a beam misalignment occurs, namely, the MU receives from the serving BS via a side lobe. We assume, as in [9], that the probability p_{bm}^M that there is a beam misalignment event for an MU moving at velocity v is equal to:

$$p_{\text{bm}}^M(v) = 1 - \exp(-v\nu_{M,b}\tau), \quad (2)$$

where $\nu_{M,b}$ is the time intensity of beam reselections. Hence, taking this beam misalignment into account, the antenna gain $g_{M,0}$ at the serving BS of the typical MU is given by

$$g_{M,0} = \begin{cases} G_m^M & \text{w.p. } 1 - p_{\text{bm}}^M(v) \\ G_s^M & \text{w.p. } p_{\text{bm}}^M(v). \end{cases} \quad (3)$$

D. Access policy and SINR

The micro and macro tiers share the same spectrum, so they interfere with one another. We call the tier with which the typical MU is associated the *association tier*, and the other tier, the *interferer tier*. An MU in the network always connects to the closest BS in its association tier. This BS association results in cells of BSs forming a Poisson-Voronoi (PV) tessellation [4].

Let $\ell(x) \triangleq Kx^{-\alpha}$, with $\alpha > 2$ and $K = \left(\frac{c}{4\pi f_c}\right)^2$, be the path-loss function for the system, where f_c is the carrier frequency and c is the speed of light. We assume Rayleigh fading with mean 1 between the BS located at x and the typical MU at the origin, denoted by h_x . Let P_M and P_μ denote the transmit powers for each tier, with $P_\mu < P_M$, and G_x^M and G_y^μ the antenna gains for macro and micro BSs located at x and y , respectively. We define the interference experienced by the typical MU from macro tier BSs as:

$$I_M(\Phi_M) = \sum_{x \in \Phi_M} h_x g_{M,x} P_M \ell(\|x\|).$$

Let $X_{M,0}$ be the closest macro BS to the typical MU, and $X_{\mu,0}$ be its micro counterpart. Conditioned on that the closest macro BS to which the typical MU connects is at distance $r > 0$, the SINR at the typical MU, denoted as SINR_M is expressed as

$$\text{SINR}_M = \frac{h_0 P_M g_{M,0} K r^{-\alpha}}{\sigma^2 + I_M(\Phi_M \setminus \{X_{M,0}\}) + I_\mu(\Phi_\mu)}, \quad (4)$$

where $g_{M,0}$ (given by (3)) is the gain of the serving BS of the typical MU and σ^2 is the thermal noise density.

E. Parameters

We consider that both tiers use a carrier frequency f_c in the 3.5 GHz range with a bandwidth $W = 100$ MHz. The transmit powers of macro and micro BSs are assumed to be 20 W and 4 W, respectively. The antenna gain profile we consider for numerical results is that the main lobe gain is $G_m^M = n_M$ and the side lobe gain is $G_s^M = 1/n_M$, but other gain profiles can be used. The thermal noise σ^2 is -174 dBm.Hz $^{-1}$. Finally, MU velocities are sampled from an exponential distribution with scale parameter v_u . Table I gives the parameters used in the numerical results.

III. COVERAGE PROBABILITY AND SHANNON RATE

A. Coverage probability

The coverage probability is defined as the probability that the SINR received by the typical user is greater than a predefined threshold T . Conditioning on the fact that the

TABLE I
VALUES OF NETWORK PARAMETERS

Parameter	Micro tier	Macro tier
Carrier frequency (f_c)	3.5 GHz	
Bandwidth (W)	100 MHz	
Thermal noise density (σ^2)	-174 dBm.Hz $^{-1}$	
Transmit powers (P_μ, P_M)	36 dBm	43 dBm
Beam reselection time ($T_{\mu,b}, T_{M,b}$)	23 ms	
BS handover time ($T_{\mu,c}, T_{M,c}$)	43 ms	
SSB burst periodicity (τ)	20 ms	
Path-loss exponent (α)	4	
Number of beams (n_μ, n_M)	8	
MU intensity (λ_u)	1 m^{-2}	
BS intensities (λ_μ, λ_M)	0.1 m^{-2}	0.02 m^{-2}
Velocity scale parameter (v_u)	$[1, 10, 35] \text{ m.s}^{-1}$	
Maximum SINR (Q_{\max})	30 dB	

typical MU is associated with a macro BS, the coverage probability p_M can be expressed as follows.

Theorem 1 (Coverage probability): The coverage probability with the macro tier association is

$$p_M(v, T) = (1 - p_{\text{bm}}^M(v))q_M(G_m^M, T) + p_{\text{bm}}^M(v)q_M(G_s^M, T), \quad (5)$$

where

$$q_M(G, T) = \pi \lambda_M \int_{r \geq 0} e^{-\pi \lambda_M r} \exp\left(-\frac{T \sigma^2}{P_M K G} r^{1/\delta}\right) \exp\left(-\pi r \left(\frac{T}{P_M G}\right)^\delta (\lambda_M P_M^\delta \rho_M(G, T) + \lambda_\mu P_\mu^\delta \kappa_\mu)\right) dr.$$

Here $\delta \triangleq 2/\alpha$, and ρ_M and κ_μ are given by

$$\begin{aligned} \rho_M(G, T) &= p_{M,m}(G_m^M)^\delta \int_{\left(\frac{T G_m^M}{G}\right)^\delta}^\infty \frac{du}{1 + u^{1/\delta}} \\ &+ (1 - p_{M,m})(G_s^M)^\delta \int_{\left(\frac{T G_s^M}{G}\right)^\delta}^\infty \frac{du}{1 + u^{1/\delta}}, \\ \kappa_\mu &= \left(p_{\mu,m}(G_m^\mu)^\delta + (1 - p_{\mu,m})(G_s^\mu)^\delta\right) \int_0^\infty \frac{du}{1 + u^{1/\delta}}. \end{aligned}$$

The coverage probability in the micro tier is obtained by swapping the role of both tiers.

Proof: The proof is given in Appendix A. ■

In an interference-limited network, i.e., when $\sigma^2 = 0$, the functions q_M and q_μ simplify to

$$q_M(G, T) = \frac{1}{1 + \left(\frac{T}{G}\right)^\delta (\rho_M(G, T) + \Omega \kappa_\mu)} \quad (6)$$

$$q_\mu(G, T) = \frac{1}{1 + \left(\frac{T}{G}\right)^\delta (\rho_\mu(G, T) + \frac{1}{\Omega} \kappa_M)}, \quad (7)$$

where $\Omega = \frac{\lambda_\mu P_\mu^\delta}{\lambda_M P_M^\delta}$. When setting $\Omega = 0$, which corresponds to setting either $\lambda_\mu = 0$ or $P_\mu = 0$, i.e., removing the micro tier, we obtain the same formula as the one from [5].

B. Shannon rate

We denote the Shannon rate per Hertz in each tier by $\mathcal{R}_M = \mathbb{E}[\log(1 + \text{SINR}_M)]$ and $\mathcal{R}_\mu = \mathbb{E}[\log(1 + \text{SINR}_\mu)]$. Due to

RF imperfections and modulation schemes, we set Q_{\max} to be the maximum achievable SINR. Thus, we have

$$\mathcal{R}_M = \int_0^{Q_{\max}} \frac{p_M(T)}{1 + T} dT \quad (8)$$

in the macro tier. The same applies in the micro tier by using the appropriate values. Using Theorem 1, the Shannon rate achieved by the typical MU can be expressed as

$$\mathcal{R}_M(v) = (1 - p_{\text{bm}}^M(v))\mathcal{R}_{M,m} + p_{\text{bm}}^M(v)\mathcal{R}_{M,s}, \quad (9)$$

where $\mathcal{R}_{M,m}$ is the Shannon rate achieved with no beam misalignment, while $\mathcal{R}_{M,s}$ is its equivalent with beam misalignment (as discussed in Section II-C).

Notice that, from (8), comparing the Shannon rates in each tier corresponds to comparing the respective coverage probabilities. The term Ω in (7) plays an important role in determining the values of the Shannon rates in each tier. If the antenna gain profile is the same in both tiers, and if $\Omega > 1$, i.e., $\lambda_\mu P_\mu^\delta > \lambda_M P_M^\delta$, the Shannon rate in the micro tier is always larger than that in the macro tier. In the rest of the paper, we assume that we are in this setup.

Equation (8) gives the Shannon rate for the typical MU without considering the overheads associated with beam reselections during intra-cell mobility and BS handovers during inter-cell mobility. To take these overheads into account, we now define the effective Shannon rate.

C. Effective Shannon rate (ESR)

When an MU crosses the boundary of the cell of its serving BS, it performs a handover to the next BS. Beam reselections also happen inside the cell when the MU crosses the boundary between two beams. The time intensity $\nu_{M,c}$ of BS handovers and the time intensity $\nu_{M,b}$ of beam reselections are given in Theorem 2 of [9]:

$$\nu_{M,c} = \frac{4\sqrt{\lambda_M}}{\pi} v, \quad \nu_{M,b} = \frac{n_M \sqrt{\lambda_M}}{\pi} v.$$

When we take beam misalignment into account, if the time intensity of beam reselections is higher than the frequency $1/\tau$ of SSB bursts, beam reselections happen during each SSB burst. If the time intensity of beam reselections is lower, the user may stay in the same reference beam between two consecutive SSB bursts. Thus the effective time intensity $\nu_{M,e}$ of beam reselections in macro tier is

$$\nu_{M,e} = \min\left(\frac{1}{\tau}, \nu_{M,b}\right).$$

Let $T_{M,c}$ and $T_{M,b}$ be the time of a BS handover and a beam reselection, respectively, in the macro tier. Then we can express the total overhead per unit of time T_o^M as

$$T_o^M = \nu_{M,c} T_{M,c} + \nu_{M,e} T_{M,b}.$$

Using the previous definitions and results, we define the ESR $\mathcal{R}_{M,\text{eff}}$ in the macro tier as

$$\mathcal{R}_{M,\text{eff}}(v) = \mathcal{R}_M (1 - T_o^M(v))^+, \quad (10)$$

where $x^+ \triangleq \max(0, x)$.

IV. VELOCITY-BASED ASSOCIATION POLICIES IN A SIMPLE CASE

In this section, we introduce the notion of velocity-based association policy in a simple case with exactly one MU per cell. The notions introduced in this section will be extended in the next section to the situation where the density of MUs and the resulting resource sharing are taken into account. The results of this section are nevertheless of independent interest as they allow one to analyze the peak rate that an MU would get in an ideal network with one MU per cell and hence without bandwidth sharing between MUs of the same cell.

Here and in what follows, to decide what tier an MU should associate with, we consider *velocity-based association policies*, which are defined as follows: let $\mathcal{P}(v)$ be a criterion based on velocity (namely a subset of the positive real line). Each MU associates either with the closest macro BS or to the closest micro BS. If the MU meets the criterion $\mathcal{P}(v)$, it associates with the micro BS, and if not, to the macro BS.

A. The ESR under a given association policy

The metric we use to compare association policies is the ESR received by the typical MU as given by (10).

Lemma 1: For all velocity-based association policies \mathcal{P} , the average ESR received by the typical MU is given by

$$\mathcal{R}(\mathcal{P}) = \int_0^\infty (\mathcal{R}_{\mu,\text{eff}}(v)\mathbb{1}_{\mathcal{P}(v)} + \mathcal{R}_{M,\text{eff}}(v)\mathbb{1}_{\bar{\mathcal{P}}(v)})f(v)dv, \quad (11)$$

where $\mathbb{1}_{(\cdot)}$ is the indicator function and $\bar{\mathcal{P}}$ the negation of \mathcal{P} . This result comes from ergodic theory: we compute the spatial average of the ESR over a large ball of radius $R > 0$ centered at the origin. Using Birkhoff's ergodic theorem [10], we can prove that this spatial average converges to a constant, $\mathcal{R}(\mathcal{P})$, which is also equal to the ESR experienced by the typical MU.

B. Threshold association policies

A natural set of policies to explore is that of threshold policies: we select a velocity threshold v_T ; all MUs with velocity $v < v_T$ associate with micro BSs and those with velocity $v \geq v_T$ with macro BSs. Under such a threshold policy, the average ESR is given by

$$\mathcal{R}(v_T) = \int_0^{v_T} \mathcal{R}_{\mu,\text{eff}}(v)f(v)dv + \int_{v_T}^\infty \mathcal{R}_{M,\text{eff}}(v)f(v)dv. \quad (12)$$

We have the following Theorem:

Theorem 2 (Threshold velocity-based association policy for the one MU per cell case): Under the foregoing assumptions, there exists a unique threshold maximizing the average ESR per user. This optimal threshold v_T^* does not depend on the velocity distribution f of MUs, and is the unique solution to the following equation:

$$\mathcal{R}_{\mu,\text{eff}}(v_T^*) = \mathcal{R}_{M,\text{eff}}(v_T^*). \quad (13)$$

The proof for this Theorem is available in Appendix B. Fig. 1 shows the normalized ESR $\mathcal{R}_{M,\text{eff}}/\|\mathcal{R}_{M,\text{eff}}\|_\infty$, where $\|\cdot\|_\infty$ is the supremum norm for three exponential velocity

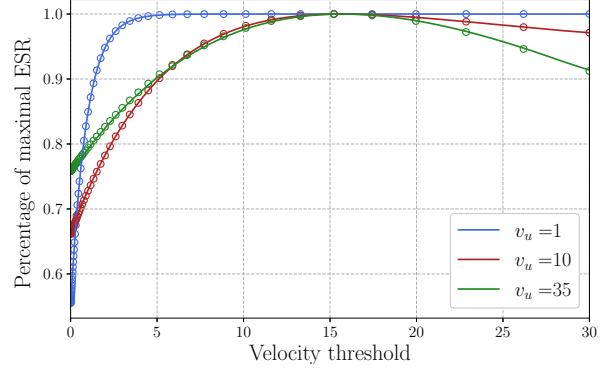


Fig. 1. Normalized ESR in the network as a function of the velocity threshold v_T . The optimal threshold does not depend on the scale parameter v_u .

distributions; we see that the maximum ESR is obtained at the same value for all velocity distributions.

This network setup mimics an ideal situation, corresponding to the peak rate that an MU receives in the network. A more realistic setup with multiple MUs is studied in the next section.

V. LOAD-DEPENDENT VELOCITY-BASED ASSOCIATION POLICY

In this section, we address the load sharing question between all MUs of the same cell. We assume a time-division multiple access (TDMA) setup where radio resources are equally shared among all MUs connecting to the same BS.

Let \mathcal{P} be a given velocity-based association policy. Under the association policy \mathcal{P} , let $Z_\mu^0(\mathcal{P})$ denote the number of users associated with the micro BS corresponding to the 0-cell, i.e., the Voronoi cell containing the typical MU located at the origin 0, and $Z_M^0(\bar{\mathcal{P}})$ be its macro counterpart. Note that $Z_\mu^0(\mathcal{P})$ and $Z_M^0(\bar{\mathcal{P}})$ are random variables due to the random network geometry and velocity. For this setup, we can modify (11) to rewrite the average ESR $\mathcal{R}_{\text{load}}$ as follows:

$$\mathcal{R}_{\text{load}}(\mathcal{P}) = \mathbb{E} \left[\frac{\mathcal{R}_{\mu,\text{eff}}(v)}{Z_\mu^0(\mathcal{P})} \mathbb{1}_{\mathcal{P}(v)} + \frac{\mathcal{R}_{M,\text{eff}}(v)}{Z_M^0(\bar{\mathcal{P}})} \mathbb{1}_{\bar{\mathcal{P}}(v)} \right]. \quad (14)$$

In (14), the number of MUs associated to the 0-cell and the ESR both depend on the geometry of each PV tessellation. We use the following heuristic to approximate $\mathcal{R}_{\text{load}}$:

$$\hat{\mathcal{R}}_{\text{load}}(\mathcal{P}) = \mathbb{E} \left[\frac{1}{Z_\mu^0(\mathcal{P})} \right] \mathbb{E} [\mathcal{R}_{\mu,\text{eff}}(v)\mathbb{1}_{\mathcal{P}(v)}] + \mathbb{E} \left[\frac{1}{Z_M^0(\bar{\mathcal{P}})} \right] \mathbb{E} [\mathcal{R}_{M,\text{eff}}(v)\mathbb{1}_{\bar{\mathcal{P}}(v)}], \quad (15)$$

where $\mathbb{E} \left[\frac{1}{Z_\mu^0(\mathcal{P})} \right]$ and $\mathbb{E} \left[\frac{1}{Z_M^0(\bar{\mathcal{P}})} \right]$ are the average inverse load in the 0-cell of micro and macro tiers, respectively.

A. Load-dependent threshold association policies

Like in Section IV-B, threshold policies remain good candidates to maximize the average ESR experienced by the typical MU. Let v_T be a velocity threshold. Under the threshold policy with threshold v_T , the intensity of the PPP of users associated

TABLE II
NUMERICAL ILLUSTRATION OF THE LOAD-THRESHOLD HEURISTIC

v_u (m.s ⁻¹)	1	10	35
v_{LT} (m.s ⁻¹)	1.30	4.90	9.41
$\mathcal{R}_{load}(v_{LT})$	0.294	0.242	0.201
Optimal ESR	0.289	0.252	0.195

with the micro tier is equal to $\lambda_u F(v_T)$, with F being the CDF of the velocity distribution, and the intensity of the PPP of users associated with the macro tier is equal to $\lambda_u(1 - F(v_T))$. Although we do not have an analytical formula for the average inverse load in the network, we can use the following heuristic:

Lemma 2: Let Φ and Ψ be two PPPs of respective intensities λ and ν . Let V_0 be the 0-cell of the PV tessellation associated with the process Φ and let $Z = 1 + |\Psi \cap V_0|$. Finally, let $L(x) = x \left(1 - \left(\frac{1}{1 + \frac{2}{7x}}\right)^{7/2}\right)$. The moment of order -1 of Z can be approximated by:

$$\mathbb{E} \left[\frac{1}{Z} \right] \approx L \left(\frac{\lambda}{\nu} \right).$$

Proof: The proof is presented in Appendix C. It uses the results from [11] and a heuristic for the distribution of the size of the typical cell of a PV tessellation developed in [12]. ■ Using this, we set the following notation:

$$\mathbb{E} \left[\frac{1}{Z_\mu^0(v_T)} \right] = L \left(\frac{\lambda_\mu}{\lambda_u F(v_T)} \right) \equiv L_\mu(v_T)$$

$$\mathbb{E} \left[\frac{1}{Z_M^0(v_T)} \right] = L \left(\frac{\lambda_M}{\lambda_u(1 - F(v_T))} \right) \equiv L_M(v_T).$$

Under a velocity-based threshold policy, the average ESR in the TDMA setup becomes

$$\hat{\mathcal{R}}_{load}(v_T) = L_\mu(v_T) \int_0^{v_T} \mathcal{R}_{\mu,eff}(v) f(v) dv + L_M(v_T) \int_{v_T}^\infty \mathcal{R}_{M,eff}(v) f(v) dv. \quad (16)$$

We give a *load-threshold heuristic* for the optimal threshold policy in the TDMA setup as follows:

$$v_{LT} \triangleq \arg \max \hat{\mathcal{R}}_{load}(v_T). \quad (17)$$

As shown in Fig. 2, we numerically observe that v_{LT} is uniquely defined, meaning that the function $\hat{\mathcal{R}}_{load}(v_T)$ goes through a unique maximum. Specifically, Fig. 2 compares the simulations under TDMA and the heuristic from (17). The difference is due to the fact that the inverse of the load and the ESR are not independent. When comparing the values of the optimal achieved ESR in simulations with the ESR achieved using the heuristic from (17) (see Table II), we can see that the obtained values do not differ significantly, which justifies the use of the heuristic.

VI. DISCUSSIONS

A. Comparison with a classical association policy

We now compare the optimal average ESR using the load-threshold heuristic based on (17) to a classical association policy based on the *maximum average received power* (MARP),

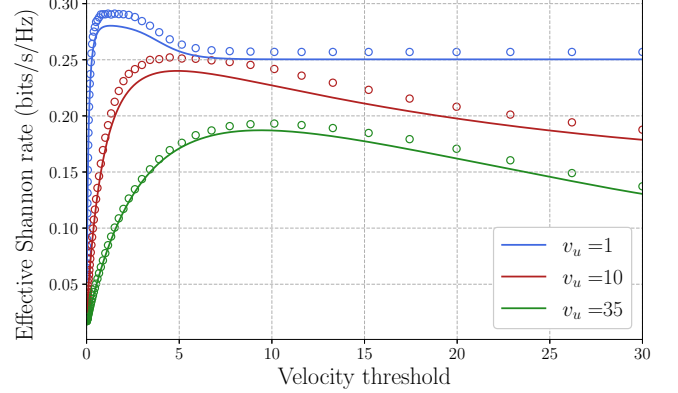


Fig. 2. Comparison of the heuristic in (16) (in plain line) and simulations (circles)

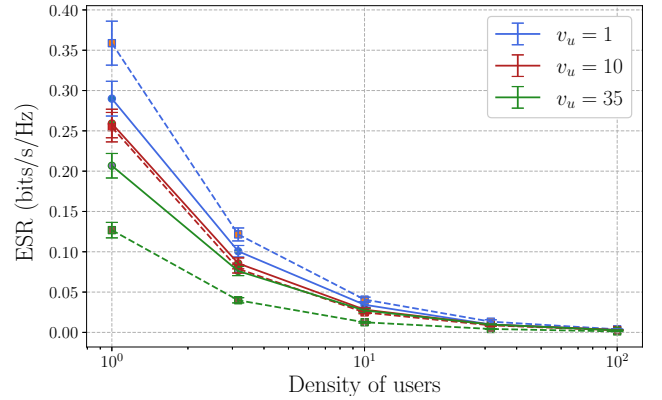


Fig. 3. Comparison between the MARP policy (square markers, dashed line) and the threshold policy under v_{LT} (circles, plain line) with 95% confidence intervals for three different velocity distributions.

where MUs connect to the BS with the highest average received power. Note that the MARP policy is primarily based on the received *power* and the threshold policy on *velocity*. Hence, this comparison allows us a fine-grained view at the ESR, which has two main components as seen from (10): 1) \mathcal{R}_M , the Shannon rate as a function of the SINR and hence the received power, and 2) $T_o^M(v)$, the time overhead as a function of the velocity v of an MU.

As shown in Fig. 3, for three different velocity distributions, while the average ESR is larger under the MARP policy at low velocities and sparsely populated networks (i.e., small λ_u), the threshold association policies perform better for higher velocities and densely populated networks (i.e., large λ_u). Thus, the velocity-based threshold policies do a better job when it comes to load balancing in a densely populated network with fast-moving MUs. In fact, for $\lambda_u = 1$ and $v_u = 35$, the threshold policies improve on the classical MARP policy with a gain of about 60%. The fact that these results are obtained for parameters representing accurately

VII. CONCLUDING REMARKS

This paper discusses a simple velocity-based association policy for multi-tier 5G networks. The proposed framework captures essential features of 5G networks, namely, beam-forming, beam misalignment, delay overheads associated with beam reselections and base station handovers, speed profiles of users, and network geometry. In this association policy, users with a velocity smaller than a threshold connect to the micro tier, otherwise to the macro tier. The threshold is optimized to maximize the effective Shannon rate that the typical user gets. The comparison of this policy with the well-known association policy based on the maximum average received power reveals that, for parameters that are typical to sub-6 GHz 5G networks, when the network is populated with users moving with variable enough speed patterns, the former outperforms the latter in terms of effective Shannon rate. This observation highlights a key tradeoff between the power (affecting the SINR) and the velocity (affecting the delay overheads). The fact that simple velocity-based schemes may significantly outperform the state of the art association policies opens an interesting line of thoughts aiming at combining such velocity-based schemes with those based on maximal received powers to provide a new and better association policy for mobile users in this 5G context.

REFERENCES

- [1] H. Holma, A. Toskala, , and T. Nakamura, *5G Technology: 3GPP New Radio*. John Wiley & Sons Ltd., 2020.
- [2] M. Giordani, M. Polese, A. Roy, D. Castor, and M. Zorzi, "A tutorial on beam management for 3GPP NR at mmWave frequencies," *IEEE Communications Surveys and Tutorials*, vol. 21, no. 1, pp. 173–196, Q1 2019.
- [3] H. Dhillon, R. K. Ganti, F. Baccelli, and J. G. Andrews, "Modeling and analysis of K -tier downlink heterogeneous cellular networks," *IEEE Journal on Selected Areas in Communications*, vol. 30, no. 3, pp. 550–560, 2012.
- [4] F. Baccelli and B. Błaszczyszyn, *Stochastic Geometry and Wireless Networks*. NoW Publishers, 2009.
- [5] J. G. Andrews, F. Baccelli, and R. K. Ganti, "A tractable approach to coverage and rate in cellular networks," *IEEE Transactions on Communications*, vol. 50, no. 11, pp. 3122–3134, 2011.
- [6] W. Bao and B. Liang, "Stochastic geometric analysis of user mobility in heterogeneous wireless networks," *IEEE Journal on Selected Areas in Communications*, vol. 33, no. 10, pp. 2212–2225, October 2015.
- [7] R. Arshad, H. ElSawy, S. Sorour, T. Y. Al-Naffouri, and M. Alouini, "Velocity-aware handover management in two-tier cellular networks," *IEEE Transactions on Wireless Communication*, vol. 16, no. 3, pp. 1851–1867, March 2017.
- [8] A. Chattopadhyay, B. Błaszczyszyn, and E. Altman, "Two-tier cellular networks for throughput maximization of static and mobile users," *IEEE Transactions on Wireless Communication*, vol. 18, no. 2, pp. 997–1010, February 2019.
- [9] S. S. Kalamkar, F. M. Abinader Jr, F. Baccelli, A. S. Marciano Fani, and L. G. Uzeda Garcia, "Stochastic geometry-based modeling and analysis of beam management in 5G," *IEEE Globecom*, 2020.
- [10] G. D. Birkhoff, "Proof of the ergodic theorem," *Proceedings of the National Academy of Sciences*, vol. 17, no. 12, pp. 656–660, 1931.
- [11] J. Mecke, "On the relationship between the 0-cell and the typical cell of a stationary random tessellation," *Pattern Recognition*, vol. 32, pp. 1645–1648, 1998.
- [12] J.-S. Ferenc and Z. Nédá, "On the size distribution of Poisson voronoi cells," *Physica A: Statistical Mechanics and its Applications*, vol. 385, no. 2, pp. 518–526, 2007.

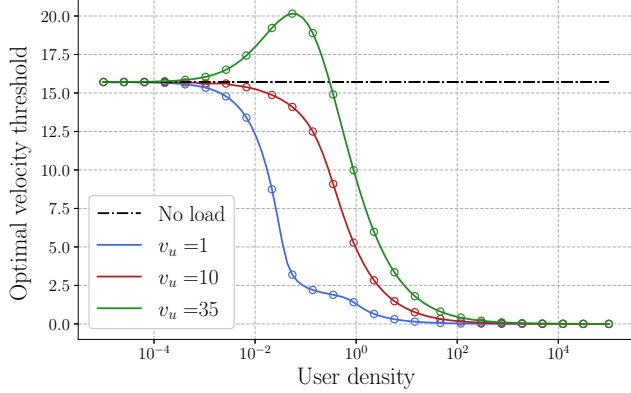


Fig. 4. Optimal velocity threshold obtained for the heuristic from (17) (in plain line) for three velocity distributions against the density of users. In black, dashed, the equivalent network without load.

the sub-6 GHz setting of 5G new radio (NR) supports the importance of the study of velocity-based threshold policies.

We now discuss the reasons behind why these policies outperform each other in certain regimes. When the MUs move slowly and the network is sparsely populated, the effects of mobility-induced beam misalignment and overheads due to beam reselections and BS handovers are negligible. Thus, the factor of better average received power outweighs the time overheads. In this case, the MARP policy outperforms the threshold policy. On the other hand, when the MUs have a higher velocity in the network that is more densely populated, the penalty due to the mobility of MUs (BS handovers, beam reselections, and beam misalignment) has more impact on the performance compared to the received power. Thus, it is beneficial to associate users based on their velocities.

B. Effect of the density of MUs in the network

The density λ_u of MUs allows us to understand how well the network performs under different load conditions. Specifically, as $\lambda_u \rightarrow 0$, the ESR in the network becomes equal to the Shannon rate without TDMA, which is its peak rate. This translates into v_{LT} (given in (17)) becoming equal to v_T^* from Theorem 2. Conversely, as $\lambda_u \rightarrow \infty$, the number of MUs associated with the micro tier becomes arbitrarily large. Because its contribution towards the ESR is larger than that from the macro tier, to maximize the average ESR, the optimal velocity threshold approaches zero.

The behavior of the optimal threshold for intermediate values of λ_u depends on the velocity distribution. Fig. 4 gives the optimal velocity threshold for the heuristic given in (17). The MU velocity is a limiting factor in the network: if the velocity of MUs is higher, the average overhead per unit of time in the macro tier increases, leading to a decrease in the contribution of the macro tier to the average ESR. To balance this effect, more MUs are associated with the micro tier, resulting in an increase in the optimal velocity threshold.

APPENDIX A: PROOF OF THEOREM 1

BSs in the macro tier are distributed according a PPP. The probability that a BS X is located at a distance $r > 0$ from the origin and the associated distribution function are equal to

$$F_M(r) = \mathbb{P}[\|X\| \leq r] = 1 - e^{-\lambda_M \pi r^2}$$

$$f_M(r) \equiv \frac{d}{dr} F_M(r) = 2\pi r e^{-\lambda_M \pi r^2}.$$

We assume that the typical MU is aligned with its the serving BS, located at distance $r > 0$ from the origin. We get:

$$\begin{aligned} \mathbb{P}[\text{SINR}_M > T | r, g_{M,0} = G_m^M] \\ &= \mathbb{P}\left[h_0 > \frac{Tr^\alpha}{P_M G_m^M} (\sigma^2 + I_M(\Phi_M \setminus \{X_{M,0}\}) + I_\mu(\Phi_\mu)) | r\right] \\ &= e^{-\frac{Tr^\alpha}{P_M G_m^M}} \mathbb{E}\left[\exp\left(-\frac{\mu T}{P_M G_m^M} I_M(\Phi_M \setminus \{X_{M,0}\})\right)\right] \times \\ &\quad \mathbb{E}\left[\exp\left(-\frac{\mu T}{P_M G_m^M} I_\mu(\Phi_\mu)\right)\right]. \end{aligned}$$

The first expectation yields

$$\begin{aligned} \mathbb{E}\left[\exp\left(-\frac{\mu T}{P_M G_m^M} I_M(\Phi_M \setminus \{X_{M,0}\})\right)\right] \\ &= \mathbb{E}\left[\prod_{x \in \Phi^M \setminus \{X\}} \mathbb{E}\left[\exp\left(-\frac{\mu T g_{M,x}}{G_m^M} r^\alpha h_x \ell(|x|)\right)\right]\right] \\ &= \mathbb{E}\left[\prod_{x \in \Phi^M \setminus \{X\}} \mathbb{E}\left[\frac{1}{1 + \frac{T g_{M,x}}{G_m^M} r^\alpha \ell(|x|)}\right]\right] \\ &= \mathbb{E}\left[\prod_{x \in \Phi^M \setminus \{X\}} \frac{p_{M,m}}{1 + \frac{T G_m^M}{G_m^M} r^\alpha \ell(|x|)} + \frac{1 - p_{M,m}}{1 + \frac{T G_m^M}{G_m^M} r^\alpha \ell(|x|)}\right]. \end{aligned}$$

Using the formula for the Laplace transform of the PPP, we have

$$\begin{aligned} \mathbb{E}\left[\exp\left(-\frac{\mu T}{P_M G_m^M} I_M(\Phi_M \setminus \{X_{M,0}\})\right)\right] &= \exp(-2\pi\lambda_M \times \\ &\quad \int_r^\infty \left(1 - \frac{p_{M,m}}{1 + \frac{T G_m^M}{G_m^M} r^\alpha \ell(|x|)} - \frac{1 - p_{M,m}}{1 + \frac{T G_m^M}{G_m^M} r^\alpha \ell(|x|)}\right) z dz) \\ &\stackrel{(a)}{=} \exp\left(-\pi\lambda_M r^2 T^\delta \left(p_{M,m} \left(\frac{G_m^M}{G_m^M}\right)^\delta \int_{\left(\frac{T G_m^M}{G_m^M}\right)^{-\delta}}^\infty \frac{du}{1 + u^{1/\delta}}\right.\right. \\ &\quad \left.\left. + (1 - p_{M,m}) \left(\frac{G_m^M}{G_m^M}\right)^\delta \int_{\left(\frac{T G_m^M}{G_m^M}\right)^{-\delta}}^\infty \frac{du}{1 + u^{1/\delta}}\right)\right), \end{aligned}$$

where (a) uses the change of variables $u = \left(\frac{G_m^M}{T G_m^M}\right)^\delta \frac{z^2}{r^2}$ for the first integral and $u = \left(\frac{G_m^M}{T G_m^M}\right)^\delta \frac{z^2}{r^2}$ for the second one. Using the same steps, the second expectation yields

$$\mathbb{E}\left[\exp\left(-\frac{\mu T}{P_M G_m^M} I_\mu(\Phi_\mu)\right)\right] = \exp\left(-\pi\lambda_\mu r^2 T^\delta \left(\frac{P_\mu}{P_M G_m^M}\right)^\delta (p_{\mu,m}(G_m^M)^\delta + (1 - p_{\mu,m})(G_m^M)^\delta) C(\delta)\right),$$

where $C(\delta) \triangleq \int_0^\infty \frac{du}{1 + u^{1/\delta}} = \frac{\delta\pi}{\sin(\pi\delta)}$. We have

$$\begin{aligned} \mathbb{E}[\mathbb{P}[\text{SINR}_M > T | r, g_{M,0} = G_m^M]] \\ &= \pi\lambda_M \int_{r \geq 0} e^{-\pi\lambda_M r^2} \exp\left(-\frac{\mu T r^{1/\delta}}{P_M G_m^M}\right) \times \\ &\quad \exp\left(\left(\frac{T}{P_M G_m^M}\right)^\delta (\lambda_M P_M^\delta \rho^M(G_m^M, T) + \lambda_\mu P_\mu^\delta \kappa_\mu(\delta))\right) dr. \end{aligned}$$

Finally, we obtain

$$q_m^M(T) = \int_{r \geq 0} \mathbb{E}[\mathbb{P}[\text{SINR}_M > T | r, g_{M,0} = G_m^M]] \cdot 2\pi\lambda_M r e^{-\pi\lambda_M r^2} dr. \quad (18)$$

Making the change of variables $v = r^2$ gives the intended result for $q_m^M(T)$. To obtain the result for q_s^M , q_M^ν , and q_s^μ , we simply swap the roles of the two network tiers and use the adequate values.

APPENDIX B: PROOF OF THEOREM 2

Differentiating eq. (12) w.r.t. the velocity threshold v_T gives:

$$\frac{d}{dv_T} \mathcal{R}(v) = \mathcal{R}_{\mu, \text{eff}}(v) f(v) - \mathcal{R}_{M, \text{eff}}(v) f(v). \quad (19)$$

Granted that $f > 0$, the points where the derivative is equal to zero are solutions to:

$$\mathcal{R}_{\mu, \text{eff}}(v) = \mathcal{R}_{M, \text{eff}}(v). \quad (20)$$

Using definitions from Section III-C, we know that $\mathcal{R}_{M, \text{eff}}(v) = (1 - T_o^M(v))^+ \mathcal{R}_M$ and that $T_o^M(v) = \nu_{M,c}(v) T_{M,c} + \nu_{M,e}(v) T_{M,b}$. We also know that $\nu_{M,e}(v)$ is equal to $\nu_{M,b}(v)$ for $v < v_{M,\tau} = \frac{\pi}{n_M \tau \sqrt{\lambda_M}}$, and to $\frac{1}{\tau}$ else. We can solve for $T_o^M(v) = 1$ depending on the value of v . We obtain two solutions:

$$\begin{cases} w_{M,1} = \frac{\pi}{\sqrt{\lambda_M} (4T_{c,M} + n_M T_{b,M})} & \text{if } v < v_{M,\tau} \\ w_{M,2} = \frac{\pi}{\sqrt{\lambda_M} 4T_{c,M}} \left(\frac{T_{M,b}}{\tau} - 1\right) & \text{if } v \geq v_{M,\tau}. \end{cases}$$

Using the numerical parameters from Table I, we obtain $v_{M,\tau} = 138.84 \text{ m.s}^{-1}$, $w_{M,1} = 62.40 \text{ m.s}^{-1}$ and $w_{M,2} = 19.37 \text{ m.s}^{-1}$. Because $w_{M,2} < v_{M,\tau}$, we reject this second solution, which gives us the following formula for $\mathcal{R}_{M, \text{eff}}(v)$:

$$\mathcal{R}_{M, \text{eff}}(v) = \begin{cases} \mathcal{R}_M \left(1 - v \frac{\sqrt{\lambda_M}}{\pi} (4T_{c,M} + n_M T_{b,M})\right) & \text{for } v \leq w_{M,1} \\ 0 & \text{else} \end{cases}$$

Conducting the same analysis for the micro network, we define $v_{\mu,\tau}$, $w_{\mu,1}$ and $w_{\mu,2}$ similarly. When evaluating the numerical values, we reject the solution $w_{\mu,2}$, because $w_{\mu,2} < v_{\mu,\tau}$, and we obtain the following expression for $\mathcal{R}_{\mu,\text{eff}}$:

$$\mathcal{R}_{\mu,\text{eff}}(v) = \begin{cases} \mathcal{R}_\mu \left(1 - v \frac{\sqrt{\lambda_\mu}}{\pi} (4T_{c,\mu} + n_\mu T_{b,\mu}) \right) & \text{for } v \leq w_{\mu,1} \\ 0 & \text{else} \end{cases}$$

$\mathcal{R}_{M,\text{eff}}(v)$ and $\mathcal{R}_{\mu,\text{eff}}(v)$ are two linear functions of v , with negative slopes respectively equal to $-\mathcal{R}_M \frac{\sqrt{\lambda_M}}{\pi} (4T_{c,M} + n_M T_{b,M})$ and $-\mathcal{R}_\mu \frac{\sqrt{\lambda_\mu}}{\pi} (4T_{c,\mu} + n_\mu T_{b,\mu})$. Because $\mathcal{R}_\mu > \mathcal{R}_M$ and $\lambda_\mu > \lambda_M$, the lines intersect at a unique point, and we can conclude that (20) admits a unique positive solution, denoted as v_T^* .

Finally, because $\mathcal{R}_{\mu,\text{eff}} \geq \mathcal{R}_{M,\text{eff}}$ for $v < v_T^*$ and $\mathcal{R}_{\mu,\text{eff}} \leq \mathcal{R}_{M,\text{eff}}$ for $v > v_T^*$, the threshold v_T^* corresponds to a maximum of the function \mathcal{R} .

APPENDIX C: PROOF OF LEMMA 2

The number of users in Ψ in the zero-cell of the Poisson-Voronoi tessellation associated with Φ is distributed with a Poisson law of parameter $\nu|V_0|$, where V_0 is the 0-cell of the tessellation. We have:

$$\begin{aligned} \mathbb{E} \left[\frac{1}{1 + |\Psi(V_0)|} \right] &= \mathbb{E} \left[\sum_{N=0}^{\infty} \frac{1}{1+N} \mathbb{P}[|\Psi(V_0)| = N] \right] \\ &= \mathbb{E} \left[\sum_{N=0}^{\infty} \frac{1}{1+N} \frac{(\nu|V_0|)^N}{N!} e^{-\nu|V_0|} \right] \\ &= \mathbb{E} \left[\frac{1}{\nu|V_0|} \left(1 - e^{-\nu|V_0|} \right) \right]. \end{aligned}$$

We can use Equation (2.2) from [11], which states that for all measurable function g , we have $\mathbb{E}[g(V_0)] = \mathbb{E}[|V|]^{-1} \mathbb{E}[|V|g(V)]$, where V is the typical cell of the tessellation. Since we are in a Poisson-Voronoi setup, meaning that $\mathbb{E}[|V|] = 1/\lambda$. Then, with $g(V) = \frac{1}{V} (1 - e^{-\nu V})$,

$$\mathbb{E} \left[\frac{1}{1 + |\Psi(V_0)|} \right] = \frac{\lambda}{\nu} \left(1 - \mathbb{E} \left[e^{-\frac{\nu}{\lambda}(\lambda|V|)} \right] \right). \quad (21)$$

Equation (21) is the Laplace transform of the random variable $\lambda|V|$ at $s = \frac{\nu}{\lambda}$. [12] provides a heuristic for the distribution of this random variable, which is

$$p(y) = Ay^{5/2} \exp \left(-\frac{7}{2}y \right), \quad (22)$$

where $A = \frac{(\frac{7}{2})^{\frac{7}{2}}}{\Gamma(\frac{7}{2})}$ is a normalization constant. Using this result, the Laplace transform is approximated by:

$$\mathbb{E} \left[e^{-\nu|V|} \right] = \left(1 + \frac{2\nu}{7\lambda} \right)^{-7/2}.$$

Substituting this result in (21) leads to the result.