The final version of this paper has been accepted for publication in the
IEEE Transactions on Wireless Communications - 2015 - in press
Digital Object Identifier 10.1109/TWC.2015.2496958

Context-Aware Handover Policies in HetNets

Francesco Guidolin Irene Pappalardo Andrea Zanella Michele Zorzi
Dept. of Information Engineering, University of Padova, via Gradenigo 6/B, 35131 Padova, Italy
{fguidolin, pappalar, zanella, zorzi}@dei.unipd.it

***Abstract*—Next generation cellular systems are expected to entail a wide variety of wireless coverage zones, with cells of different sizes and capacities that can overlap in space and share the transmission resources. In this scenario, which is referred to as *Heterogeneous Networks* (HetNets), a fundamental challenge is the management of the handover process between macro, femto and pico cells. To limit the number of handovers and the signaling between the cells, it will hence be crucial to manage the users mobility considering the context parameters, such as cells size, traffic loads, and user velocity. In this paper, we propose a theoretical model to characterize the performance of a mobile user in a HetNet scenario as a function of the user's mobility, the power profile of the neighboring cells, the handover parameters and the traffic load of the different cells. We propose a Markov-based framework to model the handover process for the mobile user and we derive an optimal context-dependent handover criterion. The mathematical model is validated by means of simulations, comparing the performance of our strategy with conventional handover optimization techniques in different scenarios. Finally, we show the impact of the handover regulation on the users performance and how it is possible to improve the users capacity exploiting context information.**

***Index Terms*—Small cells, HetNets, Handover, discrete time Markov chain, context-awareness, load balancing.**

I. Introduction

Global mobile data traffic is expected to increase exponentially in the next years, reaching 15.9 exabytes per month by 2018 [1]. One of the most promising approaches to face this challenge is the so-called Heterogeneous Network (HetNet) paradigm, which basically consists in enriching the current cellular network with a number of smaller and simpler Base Stations (BSs), having widely varying transmit powers, coverage areas, carrier frequencies, types of backhaul connections and communication protocols. The deployment of pico and/or femto BSs *within* the macrocell, indeed, can provide higher connection speed and better coverage to the mobile users located at the border of the macrocell or in regions with high traffic demand.

While increasing the efficiency of the cellular networks, HetNets also raise several technical challenges related to user management [2]. An important aspect is related to the handover (HO) process of mobile users that, differently from classical cellular networks, have to deal with cells of widely varying coverage areas. In general, the HO process, standardized by the 3rd Generation Partnership Project (3GPP) [3], is triggered by the User Equipment (UE), which periodically measures the Reference Symbols Received Power (RSRP) from the surrounding cells. When the difference between the RSRP of a neighboring cell and that of the serving cell is higher than a fixed HO hysteresis value, γ_{th}, (event A3 in [4]), the HO process starts, as exemplified in Fig. 1. If this condition holds for a period of time equal to the *Time-To-Trigger* (TTT) parameter, the HO is finalized and the UE connects to the BS with the strongest RSRP.

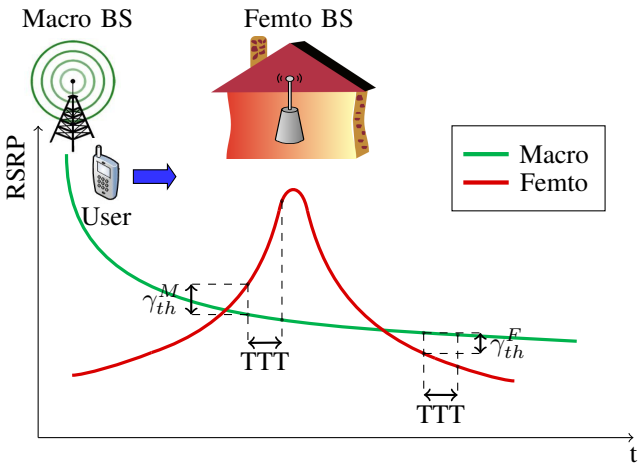

Fig. 1: Example of the decay of the power profile from the M-BS and F-BS as the UE moves away from the M-BS and towards the F-BS.

The static setting of the HO hysteresis and TTT values adopted in traditional scenarios with only macrocells is no longer effective for HetNet systems, because of the large variety in cell characteristics [5], [6]. With large values of TTT and of the hysteresis margin, the UE will likely experience a severe degradation of the RSRP during the TTT period when crossing a small cell, a problem that is generally referred to as *HO Failure*. On the other hand, short TTT and low hysteresis margin may cause *HO Ping-Pong*, i.e., frequent HOs to/from the M-BS, which yields performance losses due to signaling overhead and handover times. Reducing HO failure and ping-pong rates are clearly conflicting objectives, and the HO policy needs to trade off the two aspects [7].

Another challenge of HetNet management is the so called *Load Balancing*, which consists in mitigating congestion in cellular networks by offloading users from overloaded cells to lightly loaded neighboring cells. This problem has been mostly addressed in homogeneous networks, with only macrocells. Load Balancing in HetNets is more involved due to the disparities in cell sizes and transmit powers. In order to achieve the desired efficiency from the deployment of small cells, hence, the handover decision needs also to be load-aware. Indeed, by properly adapting the hysteresis margin, mobile

A preliminary version of this work was presented at the 13th IEEE IFIP Annual Mediterranean Ad Hoc Networking Workshop, June 2014, [8].

This work was supported by the project "A Novel Approach to Wireless Networking based on Cognitive Science and Distributed Intelligence," funded by Fondazione CaRiPaRo under the framework Progetto di Eccellenza 2012.

users may be encouraged to switch to small BSs that are lightly loaded to get higher data rates. As a consequence, macrocells will also have the possibility to better serve their remaining users.

In this paper, we make a step forward towards the design of context-aware HO policies by first presenting a theoretical model that describes the evolution of the UE state along its trajectory, within a basic but representative HetNet scenario. Second, we determine the expression of the average UE performance as a function of the HO parameters and other context parameters, such as the UE speed, the power profiles of the macro/pico/femto BSs, the cell load factors, and the channel model. The mathematical framework we developed can accommodate different performance metrics, such as the HO failure rate, the ping-pong rate, or the average Shannon capacity, which is the one actually considered in this work. The mathematical model is then used to design a context-aware HO policy (CAHP) that selects the HO parameters to maximize the performance metric with respect to the UE environment and channel conditions.

A preliminary and partial version of this work was presented in [8], where we advanced the idea of modelling the handover process by means of a non-homogeneous Markov chain, and we exploited the model to define a context-aware HO policy that was shown to improve the average performance of a mobile UE with respect to the context-agnostic policies. For the sake of completeness and self-consistency, this manuscript embodies that work that, however, is here presented in greater detail. Furthermore, upon such a base, in this paper we extend the context-aware HO policy to take into account the traffic load of the different cells. This is obtained by dynamically changing the HO hysteresis values to be considered in the HO process according to the load information broadcast by the cells, as detailed later in the manuscript. Furthermore, in Appendix B we sketch a possible generalization of the model to a HetNet scenario with multiple femtocells.

A similar work has been proposed in [9], where the authors develop a mathematical model for the HO procedure and derive a closed-form expression of the UE outage probability. Their policy selects the TTT and margin parameters in order to minimize the specific metric of handover failure rate. However, they do not consider the problem of load balancing among cells and, moreover, make the assumption that the UE trajectory with respect to the position of the BSs is known to the UE. Our work, instead, proposes a more general model, and defines a context-aware HO strategy based on the more realistic assumption that the UE's trajectory with respect to the location of the BSs is unknown and that the cells are loaded.

The rest of the paper is organized as follows. Sec. II provides an overview of related prior work. Sec. III introduces the channel propagation model and the HO mechanism, and derives the UE performance metric. Sec. IV presents the analysis of the HO process by means of a discrete time Markov chain, while Sec. V extends the previous model considering also the cell loads. The extension of the model to the multicell scenario is briefly discussed in Appendix B. Sec. VI formulates our context-aware HO optimization policy (CAHP), while Sec. VII provides some results for different scenarios, in comparison also with other standard strategies. Finally, Sec. VIII concludes the paper.

II. Prior Work

Recent surveys on self-organizing networks (SON) [10] and on mobility management in HetNets [11] clearly show that a proper configuration of the system parameters is both crucial for the overall throughput and also challenging due to the heterogeneity of the network. Some works in the literature focus on the theoretical characterization of key handover performance metrics. The authors in [14] express the relation between HO failure and ping-pong rates as a function of TTT, hysteresis margin, and user velocity. Similarly, in [15] the HO failure probability is derived as a function of the sampling period used by the user to collect the measurements from the neighboring cells, i.e., the Layer 3 filtering period. In both works however fast fading and shadowing statistics are neglected in the propagation model. In [16], instead, a closed-form expression of the HO failure rate is provided, taking into account also channel fading. The most severe limitation of the works in [14]–[16] is the assumption that small coverage areas are modeled as perfect circles that, while allowing their analytical tractability, is quite unrealistic. A study of more general user trajectories is presented in [17], where the authors propose a realistic user mobility model, and present analytic expressions for the HO rate, i.e., the expected number of HOs per unit time, and the cell sojourn time, i.e., the expected duration that the user stays within a particular serving cell.

Several solutions in the literature consider to adapt some HO parameters to the UE mobility conditions. In [12], for instance, the authors propose an algorithm that, while keeping the TTT and hysteresis margin constant, adaptively modifies the Cell Individual Offset (CIO) parameter, which is a margin to be added to the RSRP for load management purposes. The authors show that a UE can detect changes in its mobility pattern by monitoring the changes of the type of HO failure events (e.g., too early/late HO events, HO failures, or HO to the wrong cell) and, hence, can adjust the specified CIO parameter to minimize both the HO failure and the ping-pong rates.

In [13] an extensive simulation campaign is conducted in SONs to compute the Radio Link Failure (RLF)[1] rate for different UE speeds and types of handover, i.e., macro-to-macro and macro-to-pico handover. The proposed policy selects the TTT parameter that guarantees that the RLF rate is below a certain threshold. Reference [18] analyzes the Cell Range Expansion (CRE) technique that consists in enlarging the small cell coverage in order to balance the users load. The authors simulate the effect of both CRE bias and hysteresis margin on the HO failure and ping-pong rates, while fixing the TTT parameter.

A different approach is presented in [19] where the HO decision is based on a mobility prediction algorithm that estimates the residence time of the UE in the possible target cell. The proposed policy allows the UE to switch to the target cell only when the estimated residence time is above a certain

[1] According to the standard [3], a RLF is declared when the user SINR remains below a certain threshold Q_{out} for a specified amount of time (usually 1 s).

threshold. A similar procedure is considered in [20] where a mobility state estimation algorithm groups UEs into three speed classes and assigns a fixed TTT value to each of them, such that high speed UEs avoid the HO to pico cells, while lower speed UEs perform HO in order to minimize their RLF rate.

In these works, however, all users are assumed to have full access to the entire cell resources, irrespective of the current traffic load of each cell, which is unrealistic. The load balancing problem has been studied in [21], where the authors analyze the impact of the CRE parameter on the system capacity through the Cumulative Distribution Function (CDF) of the Signal to Interference plus Noise Ratio (SINR). The CRE parameter is adjusted to control the number of off-loaded users and, hence, to guarantee that the overall capacity is maximized. However, [21] assumes static users and does not take into account the handover that arises with mobile users. The algorithm in [22], instead, exploits the user mobility state and, by properly changing the users CIO parameter, reduces the congestion of overloaded cells, but without optimizing TTT and the hysteresis margin. The procedure described in [23] studies the impact of both the hysteresis margins referred to HOs to macro and small cells, assumed different in general, on the HO signaling overhead while guaranteeing the load balancing condition among users. The authors of [24], instead, propose a joint algorithm that, on the one hand, tunes TTT and the hysteresis parameters to optimize the handover performance metric (defined as a weighted sum of RLF, ping pong and handover failure) and, on the other hand, adapts the handover margin to achieve a load balancing condition.

Although these solutions improve the efficiency of HO in HetNets with respect to the standard static setting of the HO parameters, to the best of our knowledge a mathematical model that describes the HO performance as a function of the scenario parameters, such as the pathloss coefficients, the UE speed, and the cell load factors, is still lacking. In [25] we addressed this gap by proposing an approximate analytical expression for the mobile UEs performance, which is then used to define a TTT selection policy that maximizes the average Shannon capacity perceived by the UE along its trajectory. However, fading effects and load balancing conditions were not considered.

III. System Model

For the sake of simplicity, we focus on a basic scenario consisting of a macro BS (M-BS) and a femto BS (F-BS) placed at distance d_{MF}, and using the same frequency band. Despite its simplicity, this model still presents the fundamental issues related to HO in HetNets and, hence, is representative of the targeted scenario. In any case, the approach we propose in this manuscript can be generalized to more complex scenarios with multiple overlapping femtocells, though at the cost of a more involved notation and argumentation, as discussed in Appendix B.

For convenience, we define the UE's trajectory with respect to a reference circle $\mathcal{H}$ of radius R centered at the F-BS. We assume that the UE moves at constant speed v, following a straight trajectory. With reference to the polar coordinate system depicted in Fig. 2, the trajectory is then uniquely identified by the angular coordinate ϕ of point $\mathbf{b}$ where the UE crosses the border $\mathcal{H}$, and by the incidence angle ω formed by the trajectory with respect to the radius passing through $\mathbf{b}$. We assume that the UE can enter the femtocell from any point and with any angle, so that the parameters ϕ and ω are modeled as independent random variables with uniform distribution in the intervals $[0, 2\pi]$ and $[-\pi/2, \pi/2]$, respectively.

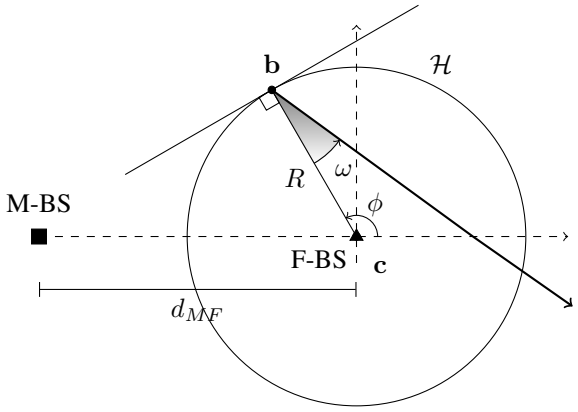

Fig. 2: Reference scenario: macrocell BS – M-BS (■), femtocell BS – F-BS (▲), and HO line $\mathcal{H}$ approximated as a circle of radius R and center $\mathbf{c}$. Linear trajectory followed by a UE when entering the femtocell at point $\mathbf{b}$ with incidence angle ω.

In the remainder of this section we describe the channel model, the HO process and the target performance metric considered in this work.

A. Propagation model

At time t, a mobile UE at position $\mathbf{a}$ measures an RSRP $\Gamma_M(\mathbf{a}, t)$ from the M-BS, and $\Gamma_F(\mathbf{a}, t)$ from the F-BS.

We assume a path-loss plus fading propagation model [26], according to which the RSRP from the h-BS, with $h \in \{M, F\}$, is given by

$$\Gamma_h(\mathbf{a}, t) = \Gamma_h^{tx}\, g_h(\mathbf{a})\, \alpha_h(t)\,, \tag{1}$$

where Γ_h^{tx} is the transmit power of the h-BS, $g_h(\mathbf{a})$ is the pathloss gain, which depends only on the distance of point $\mathbf{a}$ from the h-BS, while $\alpha_h(t)$ is the fast-fading channel gain at time t. We assume that the fading is Rayleigh distributed, i.e., $\alpha_h(t)$ is an exponential random variable with unit mean and coherence time [27]

$$T_c = \sqrt{\frac{9}{16\pi}}\frac{1}{f_d} = \sqrt{\frac{9}{16\pi}}\frac{c}{v f_c}\,, \tag{2}$$

where f_d and f_c are the Doppler and the carrier frequencies, respectively, c is the speed of light, and v is the UE's speed. Due to fading, channel fluctuations can cause the HO process to be improperly triggered, thus generating the ping-pong effect. The duration of the channel outage is a well studied metric in the literature to model this phenomenon (e.g., see [28], [29]).

Since the considered scenario is interference-limited, we can neglect the noise term and approximate the SINR $\gamma_h(\mathbf{a}, t)$ experienced by a UE connected to the h-BS at time t and in

position $\mathbf{a}$ as[2]

$$\gamma_h(\mathbf{a},t) = \bar{\gamma}_h(\mathbf{a})\xi_h(t)\,, \quad h \in \{M,F\}\,; \tag{3}$$

where

$$\bar{\gamma}_M(\mathbf{a}) = \frac{\Gamma_M^{tx}\, g_M(\mathbf{a})}{\Gamma_F^{tx}\, g_F(\mathbf{a})}\,, \quad \bar{\gamma}_F(\mathbf{a}) = \frac{\Gamma_F^{tx}\, g_F(\mathbf{a})}{\Gamma_M^{tx}\, g_M(\mathbf{a})}\,, \tag{4}$$

are the deterministic components of the SINR, while

$$\xi_M(t) = \frac{\alpha_M(t)}{\alpha_F(t)}\,, \quad \xi_F(t) = \frac{\alpha_F(t)}{\alpha_M(t)}\,, \tag{5}$$

account for the random variations due to fading.[3]

B. Handover performance model

The HO process is driven by the UE's instantaneous RSRP. If the difference between the RSRP of the serving and the target cell drops below the HO threshold γ_{th}, the TTT timer is initialized to a certain value T and the countdown starts. Whenever the RSRP difference returns above the HO threshold, however, the countdown is aborted and the HO procedure is interrupted. Conversely, if it remains below the threshold for the entire interval T, then the UE disconnects from the serving BS and connects to the new BS. This switching process takes a time T_H that accounts for the network procedures to connect the UE to the target BS. We remark here that the above condition on the RSRP difference can be translated to an equivalent condition on the SINR experienced by the UE where the power received from the target cell is the interference. Hence, we will use this latter notation in the following.

C. Mean trajectory performance

For any given point $\mathbf{a}$, we can then define the connection state S of the UE to be M, F or H depending on whether the UE is connected to the **M**-BS, the **F**-BS or is temporarily disconnected because **H**anding over from one to the other.

Given an arbitrary straight path ℓ, we define the mean trajectory performance as

$$\mathcal{C}_\ell = \frac{1}{|\ell|}\int_\ell \sum_{S\in\{M,F,H\}} C_S(\mathbf{a})\chi_{\mathbf{a}}(S)\mathrm{d}\mathbf{a}\,; \tag{6}$$

where $|\ell|$ is the trajectory's length, $\int_\ell$ is the line integral along the trajectory, $\chi_{\mathbf{a}}(S)$ is 1 if the UE's state at point $\mathbf{a}$ is S and zero otherwise, while $C_S(\mathbf{a})$ is the performance experienced by the UE at point $\mathbf{a}$ along the trajectory, given that it is in state $S \in \{M,F,H\}$.

Since the UE can follow any trajectory, we average the mean trajectory performance along all the straight lines of length L that enter the femtocell with random incidence angle, thus obtaining[4]

$$\mathcal{C}_L = \frac{1}{L\pi}\int_{-\pi/2}^{\pi/2}\int_0^L \sum_{S\in\{M,F,H\}} C_S(\mathbf{a}(x,\omega))\chi_{\mathbf{a}(x,\omega)}(S)\mathrm{d}x\,\mathrm{d}\omega\,, \tag{7}$$

with $\mathbf{a}(x,\omega)$ being the point at distance x from $\mathbf{b}$ along the trajectory with incidence angle ω.

Now, the term $\chi_{\mathbf{a}(x,\omega)}(S)$ is random, depending on the evolution of the SINR in the previous time interval of length T. Taking the expectation of (7) with respect to the random variables $\xi_h(t)$, $h \in \{M,F\}$, defined in (5), we hence get

$$\bar{\mathcal{C}}_L = \frac{1}{L\pi}\int_{-\pi/2}^{\pi/2}\int_0^L \sum_{S\in\{M,F,H\}} \bar{C}_S(\mathbf{a}(x,\omega))\mathrm{P}_S[\mathbf{a}(x,\omega)]\mathrm{d}x\,\mathrm{d}\omega\,, \tag{8}$$

where $\bar{C}_S(\mathbf{a}(x,\omega))$ is the average performance at point $\mathbf{a}(x,\omega)$, given that the UE's state at point $\mathbf{a}(x,\omega)$ is S, whose probability is

$$\mathrm{P}_S[\mathbf{a}(x,\omega)] = \mathbb{E}\left[\chi_{\mathbf{a}(x,\omega)}(S)\right]\,. \tag{9}$$

In this paper, we focus on the average Shannon capacity experienced by the UE while crossing the femtocell. Hence, for $S \in \{M,F\}$ we define

$$\begin{aligned}\bar{C}_S(\mathbf{a}) &= \mathbb{E}\left[\log_2\left(1+\gamma_S(\mathbf{a},t)\right)\right] \\ &= \log_2\left(\bar{\gamma}_S(\mathbf{a})\right)\,\frac{\bar{\gamma}_S(\mathbf{a})}{\bar{\gamma}_S(\mathbf{a})-1}\,;\end{aligned} \tag{10}$$

where the expression in the last row is derived in Appendix A. In order to account for the various costs of the handover process (energy, time, signaling, etc), we assume the UE experiences zero capacity while switching from one BS to the other, i.e., during T_H

$$\bar{C}_H(\mathbf{a}) = 0\,. \tag{11}$$

Unfortunately, the computation of (9) is very complex because of the time correlation of the fading process. To overcome this problem, we replace the continuous time model with a slotted-time model, where the UE's trajectory is observed at time epochs spaced apart by the fading coherence time T_c. In this way, at each slot we can approximately assume an independent fading value. Note that the sampling time, i.e., the slot duration, varies with the UE's speed, according to (2). Nonetheless, the distance covered by the UE in a time slot is constant and equal to

$$\Delta_c = vT_c = \sqrt{\frac{9}{16\pi}}\frac{c}{f_c}\,. \tag{12}$$

In the following, we will refer to the space interval Δ_c, which represents the spatial granularity of our model, as *space slot*.

We can then define the *average capacity* $\bar{\mathcal{C}}_L$ with respect to this sampled space as

$$\bar{\mathcal{C}}_L = \frac{1}{\pi}\int_{-\pi/2}^{\pi/2}\frac{1}{N_L}\sum_{k=1}^{N_L}\sum_{S\in\{M,F,H\}}\bar{C}_S(\mathbf{a}_k(\omega))\mathrm{P}_S[\mathbf{a}_k(\omega)]\,\mathrm{d}\omega \tag{13}$$

[2]The model can be extended to account for the interference from other cells, though for the sake of simplicity here we neglect other interference sources.

[3]In the simulations of Sec. VII, we relax the interference-limited assumption and take noise into consideration.

[4]Due to the symmetry of the problem, the entrance point $\mathbf{b}$ is irrelevant. Moreover, L is chosen to be large enough to allow the UE to be eventually connected back to the M-BS.

where

$$N_L = \left\lceil \frac{L}{\Delta_c} \right\rceil \tag{14}$$

is the total number of sample points along the trajectory, and $\mathrm{P}_S[\mathbf{a}_k(\omega)]$ is the probability that the UE is in state $S \in \{M, F, H\}$ at sample point $\mathbf{a}_k$ along its trajectory. In the next section, we describe a Markov model to compute the probabilities $\mathrm{P}_S[\mathbf{a}_k(\omega)]$.

We point out that the Markov analysis in the following section and the subsequent handover policy remain valid even with a more general propagation model than (1). The crucial aspect is that the independence of successive fading samples must be ensured by choosing a proper sampling period T_c for that channel model. The Rayleigh fading distribution used in (1) allows a semi-closed form expression for the probabilities $\mathrm{P}_S[\mathbf{a}_k(\omega)]$, whereas they can be obtained through numerical methods for any other fading distribution.

IV. Markov analysis of the HO performance

In this section we model the HO process by means of a *non homogeneous* discrete time Markov Chain (MC). To begin with, we denote by N_T and N_H the number of space slots covered by the UE in time T and T_H, respectively, i.e.,

$$N_T = \left\lceil \frac{vT}{\Delta_c} \right\rceil \ , \quad N_H = \left\lceil \frac{vT_H}{\Delta_c} \right\rceil \ . \tag{15}$$

At every step, the UE moves along its trajectory, and the SINR changes accordingly. As explained in the previous section, the HO process is started whenever the SINR drops below a certain threshold γ_{th}. We then define M_j and F_j, with $j \in \{0, \ldots, N_T\}$, as the MC state that is entered when the UE is connected to the M-BS or F-BS, respectively, and the SINR has remained below γ_{th} for j consecutive steps. Furthermore, we define H_j and $\tilde{H}_j$, $j \in \{1, \ldots, N_H\}$, as the MC states entered when the UE performs the macro-to-femto and femto-to-macro handover, respectively.

Assume that, at step k, the MC is in state M_j. In the following step, the MC evolves from M_j to M_{j+1} if $\gamma_M(\mathbf{a}_k, kT_c) < \gamma_{th}^M$, otherwise the MC returns to M_0 since the TTT counter is reset. Conversely, if the SINR remains below threshold when the MC is in state M_{N_T}, the UE starts the HO process to the F-BS and the MC enters state H_1. In the following N_H steps the MC deterministically crosses all the handover states H_j and ends up in state F_0, regardless of the channel conditions. At this point, the UE is connected to F-BS, and the evolution of the MC is conceptually identical to that seen for the M_j states.

A graphical representation of the non homogeneous discrete time MC is shown in Figure 3, with the transition probabilities that will be explained below.

A. Transition probabilities and transition matrix

The cumulative distribution function of the random variable ξ_h, given in (5) as the ratio of two independent and identically distributed exponential random variables, is equal to

$$\mathrm{P}[\xi_h \le x] = \frac{x}{x+1} \ , \quad x \in [0, +\infty] \ . \tag{16}$$

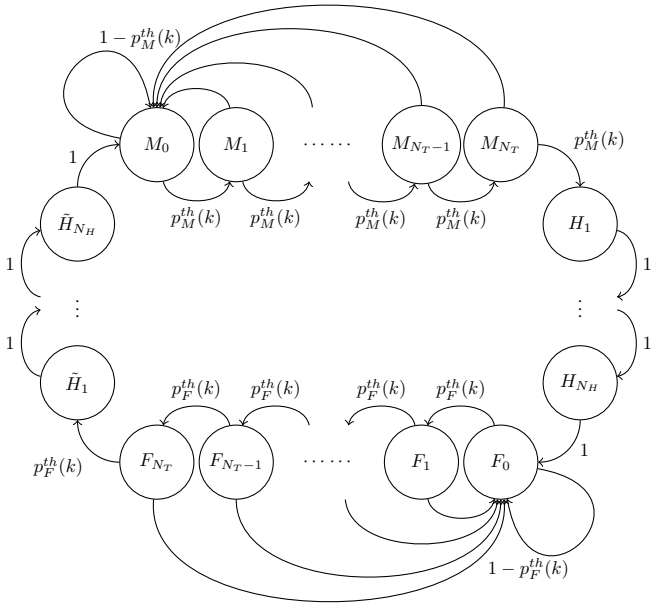

Fig. 3: Non homogeneous discrete time Markov chain referred to a scenario with arbitrary N_T and N_H. The transition probabilities are given by (17) and (18).

Using (4) and (16), the transition probability from state M_j to M_{j+1}, with $j \in \{0, \ldots, N_T\}$[5], at step k, is given by

$$p_M^{th}(k) = \mathrm{P}\left[\gamma_M(\mathbf{a}_k, kT_c) < \gamma_{th}^M\right] = \frac{\gamma_{th}^M}{\gamma_{th}^M + \bar{\gamma}_M(\mathbf{a}_k)} \, . \tag{17}$$

Similarly, the transition probability from F_j to F_{j+1} is equal to

$$p_F^{th}(k) = \mathrm{P}\left[\gamma_F(\mathbf{a}_k, kT_c) < \gamma_{th}^F\right] = \frac{\gamma_{th}^F}{\gamma_{th}^F + \bar{\gamma}_F(\mathbf{a}_k)} \, . \tag{18}$$

Note that (17) and (18) vary along the UE trajectory because of the pathloss, so that the MC is indeed non-homogeneous.

Without loss of generality, we can arrange the states according to the order $\{M_j\}$, $\{H_j\}$, $\{F_j\}$, and $\{\tilde{H}_j\}$, and in increasing order of the index j within the same set of states. The system transition matrix $\mathbf{P}(k)$ at the k-th step can then be expressed with the following sub block structure

$$\mathbf{P}(k) = \begin{bmatrix} \mathbf{M}(k) & \mathbf{V}_M^H(k) & \varnothing & \varnothing \\ \varnothing & \mathbf{H}(k) & \mathbf{V}_H^F(k) & \varnothing \\ \varnothing & \varnothing & \mathbf{F}(k) & \mathbf{V}_F^{\tilde{H}}(k) \\ \mathbf{V}_{\tilde{H}}^M(k) & \varnothing & \varnothing & \tilde{\mathbf{H}}(k) \end{bmatrix} \tag{19}$$

where the submatrices $\mathbf{M}(k)$, $\mathbf{F}(k)$, $\mathbf{H}(k)$, and $\tilde{\mathbf{H}}(k)$ are the square transition matrices within the sets $\{M_j\}$, $\{F_j\}$, $\{H_j\}$, and $\{\tilde{H}_j\}$, respectively, while $\mathbf{V}_X^Y(k)$ are the rectangular transition matrices from set X to set Y. The elements of other blocks, represented by the symbol $\varnothing$, are all equal to 0. From

[5]With $M_{N_T+1} \equiv H_1$.

the previous analysis, $\mathbf{M}(k)$ is given by

$$\mathbf{M}(k) = \begin{bmatrix} 1-p_M^{th}(k) & p_M^{th}(k) & 0 & \cdots & 0 \\ 1-p_M^{th}(k) & 0 & p_M^{th}(k) & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & 0 \\ 1-p_M^{th}(k) & 0 & 0 & \cdots & p_M^{th}(k) \\ 1-p_M^{th}(k) & 0 & 0 & \cdots & 0 \end{bmatrix} . \tag{20}$$

$\mathbf{F}(k)$ is the same as $\mathbf{M}(k)$ with $p_F^{th}(k)$ in place of $p_M^{th}(k)$, while

$$\mathbf{H}(k) = \tilde{\mathbf{H}}(k) = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \cdots & \cdots & 1 \\ 0 & 0 & \cdots & \cdots & 0 \end{bmatrix} . \tag{21}$$

Finally,

$$\mathbf{V}_H^F(k) = \mathbf{V}_{\tilde{H}}^M(k) = \begin{bmatrix} \varnothing & \varnothing \\ 1 & \varnothing \end{bmatrix} , \tag{22}$$

and

$$\mathbf{V}_M^H(k) = \begin{bmatrix} \varnothing & \varnothing \\ p_M^{th}(k) & \varnothing \end{bmatrix} , \mathbf{V}_F^{\tilde{H}}(k) = \begin{bmatrix} \varnothing & \varnothing \\ p_F^{th}(k) & \varnothing \end{bmatrix} . \tag{23}$$

The state probability vector $\mathbf{p}(k)$ at the k-th step is given by

$$\mathbf{p}(k) = \mathbf{p}(0) \prod_{i=0}^{k-1} \mathbf{P}(i) , \tag{24}$$

where $\mathbf{p}(0)$ is the state probability vector at the starting point of the UE trajectory, and $\mathbf{P}(i)$ is the transition matrix defined at the i-th step along the UE trajectory. Assuming that the UE starts its path when connected to the M-BS, we set the initial probabilities to 1 for M_0 and 0 for all the other states, so that

$$\mathbf{p}(0) = \begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix} . \tag{25}$$

We can then compute the probability that the UE is in state $S \in \{M, F, H\}$ at any given point $\mathbf{a}_k$, $k \in \{1, \ldots, N_L\}$, as the sum of the probabilities of the states $\{M_j\}$, $\{F_j\}$, and $\{H_j\} \cup \{\tilde{H}_j\}$, respectively, at step k, i.e.,

$$\mathrm{P}_S[\mathbf{a}_k] = \sum_{i \in \{S_j\}} p_i(k) , \tag{26}$$

where $p_i(k)$ is the i-th entry of the state probability vector (24).

V. Handover Decision accounting for Cell Load

In this section we consider the handover decision problem when macro and femtocells are partially loaded. In this case, handing over towards the BS with the strongest RSRP may actually yield poorer performance because of the traffic load of the new cell. As in [31], we assume that the BSs include an indication of their current traffic load in the pilot signals, so that the UEs know the average fraction of available resources for each surrounding cell. This information shall then be considered in the HO strategy, in order to select the cell with the best tradeoff between signal quality and traffic load.

Let $\lambda_S \in [0, 1]$, $S \in \{M, F\}$, denote the fraction of available resources in the cell served by S-BS. Although our model can accommodate any other scaling law, for the sake of simplicity we assume that the average performance experienced by a UE when connected to such a BS will be simply proportional to λ_S. We hence define the load-scaled average capacity of the UE in state $S \in \{M, F\}$ as follows

$$\bar{C}_S^{load}(\mathbf{a}_k) = \lambda_S \bar{C}_S(\mathbf{a}_k) = \lambda_S \log_2(\bar{\gamma}_S(\mathbf{a}_k)) \frac{\bar{\gamma}_S(\mathbf{a}_k)}{\bar{\gamma}_S(\mathbf{a}_k) - 1} , \tag{27}$$

while, as usual, we assume zero capacity during handover, i.e.,

$$\bar{C}_H^{load}(\mathbf{a}_k) = 0 . \tag{28}$$

Accordingly, the average load-scaled capacity $\bar{\mathcal{C}}_L^{load}$ along the UE trajectory is given by

$$\bar{\mathcal{C}}_L^{load} = \frac{1}{\pi} \int_{-\pi/2}^{\pi/2} \frac{1}{N_L} \sum_{k=1}^{N_L} \sum_{S \in \{M,F,H\}} \bar{C}_S^{load}(\mathbf{a}_k(\omega)) \mathrm{P}_S^{load}[\mathbf{a}_k(\omega)] \, \mathrm{d}\omega \tag{29}$$

where $\mathrm{P}_S^{load}[\mathbf{a}_k(\omega)]$ is the probability that at point $\mathbf{a}_k$ the UE is in state $S \in \{M, F, H\}$. Clearly, this probability depends on the HO policy, which shall be adjusted to account for the load conditions of the cells.

A simple way to reach this goal, with minimal impact on the HO mechanism, is to maintain the standard SINR-based HO procedure considered in the previous section, and acting on the Cell Individual Offset (CIO) of the cells, which shall be modified to account for the different traffic loads. This is equivalent to defining, for each cell S, a threshold $\gamma_{th}^{S,load}$ that depends on the current traffic loads of the macro and femtocells, respectively.

The choice of the thresholds determines the characteristics of the load-aware HO algorithm. A reasonable approach is to adapt the threshold to the cell loads in such a way that the relative performance gain experienced by the UE when changing BS is constant. Now, averaging over the fading phenomena and assuming both macro and femtocells are unloaded ($\lambda_M = \lambda_F = 1$), the HO from M-BS to F-BS is triggered when the SINR drops below the threshold γ_{th}^M. According to (10), the ratio between the average capacity of the UE in states M and F at this threshold-crossing point $\mathbf{a}_{k^*}$ is given by

$$\frac{\bar{C}_M(\mathbf{a}_{k^*})}{\bar{C}_F(\mathbf{a}_{k^*})} = \frac{\log_2(\gamma_{th}^M) \frac{\gamma_{th}^M}{\gamma_{th}^M - 1}}{\log_2(1/\gamma_{th}^M) \frac{1/\gamma_{th}^M}{1/\gamma_{th}^M - 1}} = \gamma_{th}^M , \tag{30}$$

where $\bar{\gamma}_M(\mathbf{a}_{k^*}) = \gamma_{th}^M$ and $\bar{\gamma}_F(\mathbf{a}_{k^*}) = 1/\gamma_{th}^M$. We can then set $\gamma_{th}^{M,load}$ in such a way that the ratio between the load-scaled capacities given by (27) at the new threshold-crossing point $\mathbf{a}_{k^*}^{load}$ is still equal to γ_{th}^M, i.e.,

$$\frac{\bar{C}_M^{load}(\mathbf{a}_{k^*}^{load})}{\bar{C}_F^{load}(\mathbf{a}_{k^*}^{load})} = \gamma_{th}^M . \tag{31}$$

where $\bar{\gamma}_M(\mathbf{a}_{k^*}^{load}) = \gamma_{th}^{M,load}$ and $\bar{\gamma}_F(\mathbf{a}_{k^*}^{load}) = 1/\gamma_{th}^{M,load}$. Using (27) into (31) we finally get

$$\gamma_{th}^{M,load} = \gamma_{th}^M \frac{\lambda_F}{\lambda_M} . \tag{32}$$

Repeating the same reasoning for the femto-to-macro handover, we get

$$\gamma_{th}^{F,load} = \gamma_{th}^{F} \frac{\lambda_M}{\lambda_F} \,. \tag{33}$$

Using $\gamma_{th}^{S,load}$ in place of γ_{th}^{S} in (17) and (18), we can then resort to the MC model described in the previous section to compute the average trajectory performance achieved by the load-aware HO policy. The model can then be utilized to investigate the optimal choice of the TTT parameter, as will be explained in the next section.

VI. Context-Aware HO Policy (CAHP)

The mathematical model developed in the previous sections can be used to derive a *Context-Aware HO Policy* (CAHP). The context parameters that the model is built upon consist of the transmit powers of the BSs (Γ_M^{tx} and Γ_F^{tx}), the path loss coefficients (which determine the distance-dependent path gains $g_M(\mathbf{a})$ and $g_F(\mathbf{a})$), the inter-BS distance d_{MF}, the carrier frequency f_c, and the UE speed v. In addition, the traffic load of the cells can be considered for the traffic-aware CAHP. Given these parameters, it is then possible to use the models (13) and (29) to find the value TTT that maximizes the estimated average performance experienced by the UE when crossing the area. The CAHP, hence, consists in using the optimal TTT value for the current context parameters, which are supposed to be either known by the UE or estimated from the RSRP received from the different BSs. In fact, pilot signals can carry all the necessary information, such as the pathloss exponent used in the propagation model and the cell load conditions, while the UE speed can be accurately obtained from the UE itself, with standard GPS-based systems provided by current devices.

In the remainder of this section we investigate the average UE capacity (13) when varying the context parameters, in order to gain insight on the shape of the CAHP when the cell traffic load is neglected. In the following section, we compare by simulation the performance of our CAHP against the standard handover process using static TTT values (FIX) and we extend the analysis to the model described in Sec. V, where the load of the two cells is considered.

We assume a scenario composed by a M-BS with transmission power of 46 dBm and a F-BS with transmission power of 24 dBm [30]. The BSs are placed 500 m apart. Furthermore, we set $T_H = 200$ ms, $\gamma_{th}^{M} = \gamma_{th}^{F} = 1$ dB, while T is varied with a granularity of 10 ms.

Fig. 4 shows the analytical average capacity $\bar{\mathcal{C}}_L$ given by (13) for different speeds, as a function of T. We note that the curves show a similar trend for all speed values. The sharp capacity drop for low T values is due to the ping-pong effect, which is indeed alleviated when using longer T values. In particular, the longer the channel coherence time (i.e., the lower the speed v), the larger the T required to avoid the ping pong effect, as expected. For high T values, all curves reach an asymptotic value that corresponds to the average trajectory capacity achievable when handover is not performed. The optimal T shall then trade off between the risk of ping-pong effect and the HO delay. Note that, for very high UE speeds, the maximum capacity corresponds to the asymptotic capacity. In this case, the optimal policy simply consists in always avoiding the HO, since the performance loss incurred during the HO process is never compensated by the capacity gain obtained by connecting to the F-BS.

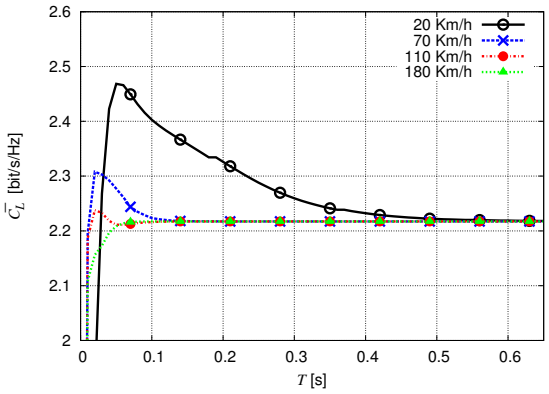

Fig. 4: Analytical average trajectory capacity obtained for different speeds, as a function of the TTT.

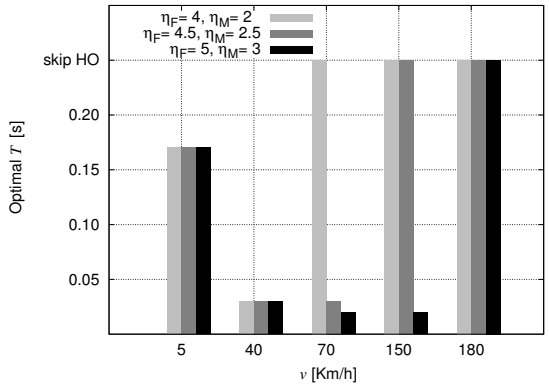

Fig. 5: Optimal T for different UE speeds v and channel parameters according to the CAHP approach.

Fig. 5 shows the optimal T values obtained from the analytical model for different speeds and scenarios. In practice, we vary the pathloss coefficients of the macro and femto BSs to change the channel profile and the femtocell coverage area, which is "small" for $\eta_F = 2, \eta_M = 4$ (radius of 9 m, left most bar), "medium", for $\eta_F = 2.5, \eta_M = 4.5$ (radius of 11 m, middle bar), and "large", for $\eta_F = 3, \eta_M = 5$ (radius of 13 m, right most bar). As predictable, the speed threshold above which the optimal policy is to skip HO depends on the femtocell range. In particular, for large cells, the losses due the HO are balanced by the higher capacity obtained by connecting to the F-BS. Therefore, skipping HO is convenient only when the UE speed is quite high. For lower speeds, instead, the optimal T is the minimum value to avoid ping-pong events due to fast fading and, hence, only depends on the channel coherence time that, in turn, depends on the UE's speed, but is independent of the size of the cells.

VII. Performance Evaluation

In this section we evaluate the performance achieved by the CAHP approach through Montecarlo simulations. In particular, we compare the mean capacity obtained by CAHP against the

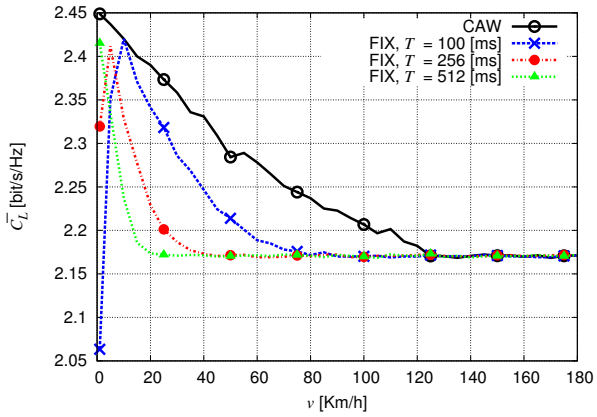

Fig. 6: Average capacity trajectory obtained with different approaches, as a function of the UE speed.

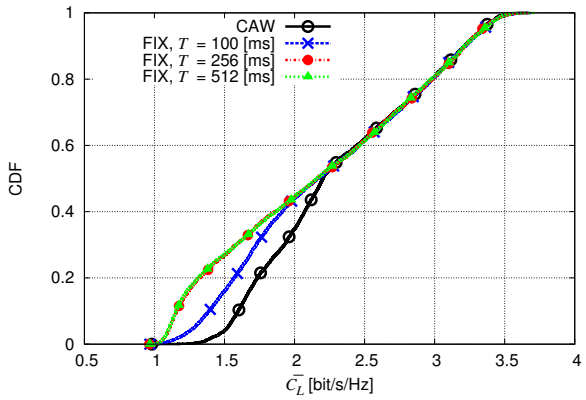

Fig. 7: Average trajectory capacity CDF for different approaches.

capacity of FIX policies that use constant TTT values, with $T \in \{100\,\text{ms}, 256\,\text{ms}, 512\,\text{ms}\}$, irrespective of the UE speed and of the other channel parameters. In the simulation we consider path loss coefficients $\eta_F = 2.5$ and $\eta_M = 4.5$ for F-BS and M-BS, respectively, the fast fading model presented in Sec. III, and a noise level equal to $\sigma^2 = -130$ dBm, obtained assuming a total downlink bandwidth of 20 MHz and a noise power spectral density of $N = k_B \mathrm{T}_0 = -143.82$ dBW/MHz, where the noise temperature T_0 is equal to 300 K and k_B is the Boltzmann constant.[6]

Fig. 6 shows the average trajectory capacity obtained in the simulations. At low speeds, the performance of the FIX policy suffers from the ping-pong effect due to low T values, while CAHP adopts a larger T that avoids HO triggering due to fast-fading fluctuations. Conversely, for higher speeds, CAHP outperforms the FIX policy by adopting sufficiently low T values to avoid the ping-pong effects, while not excessively delaying the switching to the F-BS. In particular, the higher the fixed T value, the lower the speed beyond which HO is never performed, and the higher the capacity loss compared to CAHP that, instead, performs a handover. We note that, at high speeds, all curves asymptotically converge to the same value corresponding, as in the analytical model, to the average trajectory capacity achieved when the UE remains always connected to the M-BS. The optimal HO policy consists therefore in not performing the handover to the F-BS, to avoid the loss due to two zero-capacity T_H intervals in a short time. In this case, all policies with sufficiently large T obtain the same results. Note that the asymptotic capacity given by simulations slightly differs from that given by the Markov model, as reported in Fig. 4. This small discrepancy is likely due to the simplifying assumption of the analytical model, which considers a perfectly homogeneous scenario around the femtocell center $\mathbf{c}$. The simulations, instead, consider the actual location of both BSs and the actual power received at any given point by each of them.

Fig. 7 describes the cumulative distribution function (CDF) of the average trajectory capacity for a UE speed of $v = 40$ Km/h. We note that the improvement provided by CAHP is concentrated in the lower part of the CDF. These values correspond to the trajectories that cross the femtocell area close to its center, i.e., to the location of the F-BS. In this region, a small T makes it possible to exploit the signal from the F-BS and to gain up to 50% in capacity in comparison with the case with larger T. On the contrary, the higher part of the CDF corresponds to trajectories that cross the femtocell far from the center, so that the average trajectory capacity is basically unaffected by T because HO is skipped in most cases.

[6]We verified that these results are essentially the same that would be obtained in the absence of noise.

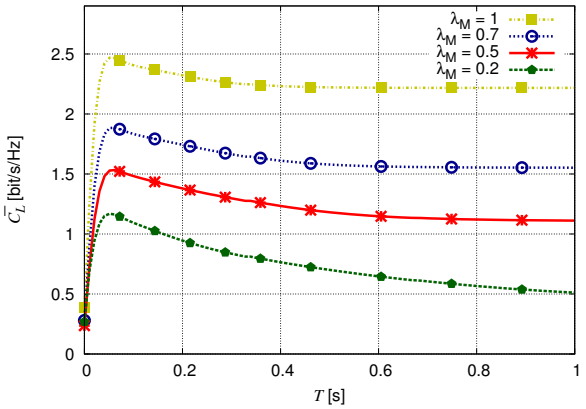

Fig. 8: Analytical average trajectory capacity obtained for different load conditions, as a function of T, with $v = 20$ Km/h.

The above results have been obtained by assuming that both the macro and the femtocell were unloaded. In the following we instead consider the case where the capacity of the cells is partially taken by other users. The pathloss coefficients from M-BS and F-BS are fixed to 4.5 and 2.5, respectively. Fig. 8 shows the analytical average trajectory capacity (29) as function of T, and with UE's speed $v = 20$ Km/h, when varying the load factor λ_M of the macrocell in the set $\lambda_M \in \{0.2, 0.5, 0.7, 1\}$, while keeping the femtocell unloaded ($\lambda_F = 1$). We can observe that the curves in Fig. 8 have the same shape, but are scaled according to λ_M. In particular, the asymptotic capacity scales proportionally to λ_M. In fact, when T is large enough, the UE does not perform any handover and remains always connected to the macrocell, and its resulting average trajectory capacity equals that of the macrocell, which is scaled by a factor λ_M with respect to the unloaded case.

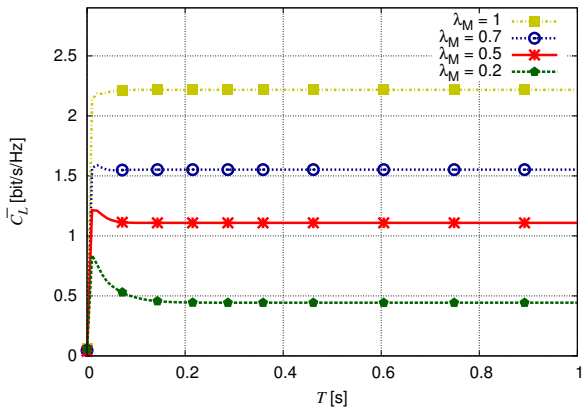

Fig. 9: Analytical average trajectory capacity obtained for different load conditions, as a function of T, with $v = 150$ Km/h.

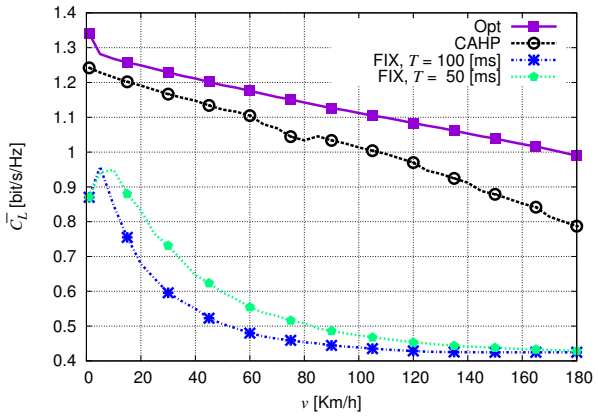

Fig. 10: Average trajectory capacity obtained with different approaches with $\lambda_M = 0.2$.

We also observe that the T value that maximizes the average trajectory capacity is the same for every load condition. The situation however changes for higher UE speed, as can be seen from Fig. 9 which reports the average capacity of the UE when varying T, with $v = 150$ Km/h. Here, CAHP encourages the UE to switch to the femtocell for highly loaded macrocells ($\lambda_M = 0.2, 0.5$), while it avoids the handover when the macrocell is unloaded. This confirms the intuition that the threshold speed increases with the load of the macrocell.

Figs. 10 and 11 show the average trajectory capacity obtained through simulations when fixing $\lambda_F = 1$ and setting λ_M equal to 0.2 and 0.7, respectively. In order to quantify the performance achieved by CAHP, we show also the capacity upper bound (Opt) computed through an exhaustive search of all possible HO policies, thus representing the best achievable performance for every user trajectory. Note that the computation of the optimal strategy requires to know in advance the fast fading gains at each point along the UE's trajectory and, hence, it is infeasible in practical scenarios. As in the previous case, we compare the performance achieved by the CAHP policy with two TTT-fixed policies, where the cell loads are not considered and T is set to 100 ms and 50 ms, respectively. As in Fig. 6, the CAHP approach achieves a substantial gain in comparison with the TTT-fixed policies for all the considered speeds. We notice that, since the capacity penalty due to T_H is larger at high speeds, the gap with the Opt policy increases with the users velocity. Moreover, the gain provided by the CAHP policy grows when the cell load is unbalanced.

This trend is further analyzed in Fig. 12. In this simulation we set $v = 60$ Km/h, while λ_M is varied from 0.1 to 1 and $\lambda_F = 1$. As expected, the average trajectory capacity increases when the macrocell is unloaded since HO is performed less frequently because the macrocell provides good enough performance. When the load at the macrocell increases, the gap between the CAHP and the TTT-fixed policies increases. The CAHP gain is due to the capability of the CAHP approach to tune the TTT considering the cell loads. In particular, when the load at the macrocell is very high, the CAHP policy achieves more than 100% performance improvement with respect to the TTT-fixed policies.

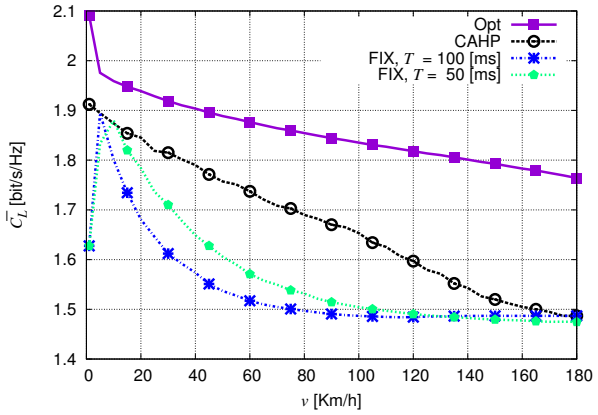

Fig. 11: Average trajectory capacity obtained with different approaches with $\lambda_M = 0.7$.

VIII. Conclusions

In this work we proposed a novel approach to optimize the handover procedure in HetNets by considering context parameters, such as the user speed, the channel gains and the load information of the cells. We derived a novel analytical framework that makes use of a Markov chain to model the evolution of the UE state during the handover process. The model was then used to derive the handover strategy that

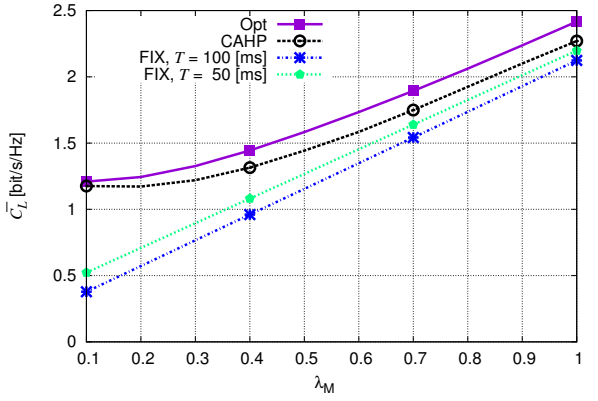

Fig. 12: Average trajectory capacity obtained with different approaches for $v = 60$ Km/h and varying λ_M from 0.1 to 1.

maximizes the UE average capacity in different scenarios, as a function of the context parameters. By adding suitable offsets to the HO thresholds, we then adjusted the mathematical model and the CAHP algorithm to account for the traffic loads of the cells. Finally, we presented a number of simulation results to assess the performance obtained by the proposed policy in comparison with standard HO policies with fixed TTT.

From this study it clearly emerges that context-awareness can indeed improve the handover process and significantly increase the performance of mobile UEs in HetNets. Although in this paper we assume that the UE trajectory is unknown, the proposed model can actually be adapted to account for exact (or statistical) knowledge of the UE path across the HetNet. In this case, the adoption of context-aware HO policies becomes even more crucial. The challenge, then, becomes the development of suitable techniques to estimate the context parameters, and the UE trajectory, in a simple and reliable manner, possibly using machine-learning approaches.

APPENDIX A

From (16), the probability density function of ξ is given by

$$f_\xi(x) = \frac{\mathrm{d}}{\mathrm{d}x}\mathrm{P}[\xi \le x] = \frac{1}{(x+1)^2}\ , \quad x \in [0, +\infty]\ . \tag{34}$$

Given $\bar{\gamma}$, the expectation of $\log_2(1+\bar{\gamma}\xi)$ is computed as

$$\begin{aligned}
\int_0^{+\infty} \log_2\left(1+\bar{\gamma}x\right) f_\xi(x)\,\mathrm{d}x &= \int_0^{+\infty} \log_2\left(1+\bar{\gamma}x\right) \frac{1}{(x+1)^2}\,\mathrm{d}x \\
&= -\beta\,\frac{\ln\left(1+\bar{\gamma}x\right)}{1+x}\bigg|_0^{+\infty} + \beta\bar{\gamma}\int_0^{+\infty} \frac{1}{1+x}\,\frac{1}{1+\bar{\gamma}x}\,\mathrm{d}x \\
&= \beta\frac{\bar{\gamma}}{\bar{\gamma}-1}\int_0^{+\infty}\left[\frac{\bar{\gamma}}{1+\bar{\gamma}x} - \frac{1}{1+x}\right]\mathrm{d}x \\
&= \frac{\bar{\gamma}}{\bar{\gamma}-1}\log_2\left(\frac{1+\bar{\gamma}x}{1+x}\right)\bigg|_0^{+\infty} \\
&= \frac{\bar{\gamma}}{\bar{\gamma}-1}\log_2\left(\bar{\gamma}\right)
\end{aligned}$$

where $\beta = \log_2 e$ and integration by parts was used to solve the integral.

APPENDIX B

We here describe a possible extension of the proposed mathematical model to a scenario with multiple femtocells. We indicate with $\mathcal{F} = \{F_1, \ldots, F_N\}$ the set of N femtocells, placed within the macrocell coverage area. At every step of its trajectory, the UE can be connected either to one of the femtocells or to the macrocell, or can be switching from the serving to the target cell. The average capacity along the whole trajectory is still computed as in (13), except for the UE state space, which is now $\{M, H\} \cup \mathcal{F}$, i.e.,

$$\bar{\mathcal{C}}_L = \frac{1}{\pi}\int_{-\pi/2}^{\pi/2}\frac{1}{N_L}\sum_{k=1}^{N_L}\sum_{S\in\{M,H\}\cup\mathcal{F}} \bar{C}_S(\mathbf{a}_k(\omega))\mathrm{P}_S[\mathbf{a}_k(\omega)]\,\mathrm{d}\omega\,. \tag{35}$$

The average capacity $\bar{C}_S(\mathbf{a}_k(\omega))$ at point $\mathbf{a}_k$ is given in (10) and (11), and the SINR $\gamma_S(\mathbf{a}_k, kT_c)$ with respect to the S-BS, $S \in \mathcal{F} \cup M$, is now given by

$$\gamma_S(\mathbf{a}_k, kT_c) = \frac{\Gamma_S(\mathbf{a}_k, kT_c)}{\sum_{S'\neq S}\Gamma_{S'}(\mathbf{a}_k, kT_c)} \tag{36}$$

where each received signal has power as in (1).

The probability $\mathrm{P}_S[\mathbf{a}_k(\omega)]$ in (35) is defined as in Sec. IV and computed from the Markov Chain described below.

The MC for the multi cell scenario is slightly more involved than the one for the single femtocell (see Fig. 3), but the principle of transition among states remains unchanged. The main difference is that we here need to take into account a TTT counter for each of the possible target BSs; the counter that expires first determines the next serving BS.

The states of the MC can be split into two classes. The first one describes the *cell states*, depicted with rectangular boxes in Fig. 13 and Fig. 14, where the UE is connected to any of the $N+1$ BSs and one or more TTTs can possibly start. We recall here that, according to the standard [3], the TTT from the UE serving cell $\mathcal{S}er$ towards the target cell $\mathcal{T}$ starts when the SINR

$$\gamma_{\mathcal{S}er,\mathcal{T}}(\mathbf{a}_k, kT_c) = \frac{\Gamma_{\mathcal{S}er}(\mathbf{a}_k, kT_c)}{\Gamma_{\mathcal{T}}(\mathbf{a}_k, kT_c)} \tag{37}$$

goes below threshold. In other words, in a multi-cell scenario the trigger condition involves the received powers of just the serving and the target BS. The cell states are defined as the $(N+1)$-tuples $< c_M, c_1, \ldots, c_N >$, where

$$c_S = \begin{cases} \mathrm{C} & \text{if } S = \mathcal{S}er \\ t & \text{otherwise}\,. \end{cases} \tag{38}$$

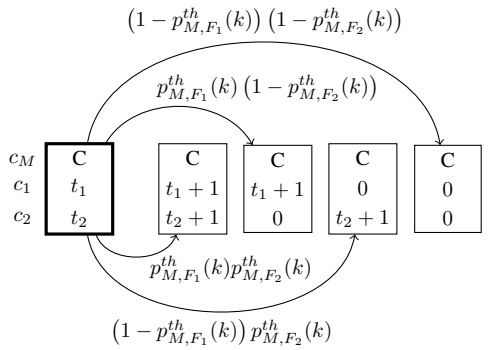

Fig. 13: Transitions from cell state $< \mathrm{C}, t_1, t_2 >$ (in bold), where $0 \le t_1, t_2 < N_T$.

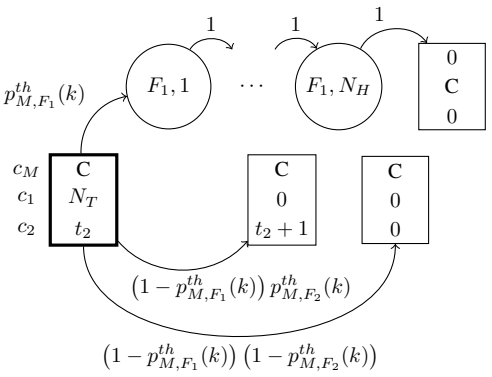

Fig. 14: Transitions from cell state $< \mathrm{C}, t_1, t_2 >$ (in bold), where $t_1 = N_T$ and $0 \le t_2 < N_T$.

The parameter C indicates the BS that the UE is currently attached to, while the number $t \in \{0, 1, \ldots, N_T\}$ indicates for how many consecutive steps the SINR $\gamma_{\mathcal{S}er,S}(\mathbf{a}_k, kT_c)$ has been below threshold, i.e., t represents the TTT counter for a possible handover to S-BS. The UE will be eventually connected to BS $S^* \neq \mathcal{S}er$ if $c_{S^*} = N_T$ and $\gamma_{\mathcal{S}er,S^*}(\mathbf{a}_k, kT_c)$ remains below threshold for one more step. Obviously, S^* is the state for which these conditions occur first.

The second class of states in the MC accounts for the handover procedures towards the new serving cell. In this case the *handover states*, depicted with circles in Fig. 14, are defined by the pair $< S, h >$ where S specifies the BS to be connected to and $h \in \{1, \ldots, N_H\}$ is the counter of the handover time.

For the sake of conciseness, we do not replicate here the rigorous analysis presented in Sec. IV for the single cell case. We prefer instead to give some intuition on how the MC evolves in this more general case.

The transitions among cell states are constrained by the fact that, if at the k-th step $c_S = t$, with $t < N_T$ and $S \neq \mathcal{S}er$, then in the following step c_S could be either $t+1$, if $\gamma_{\mathcal{S}er,S}(\mathbf{a}_k, kT_c) < \gamma_{th}^{\mathcal{S}er}$, or 0 otherwise, i.e., the counter to S-BS is reset if its SINR goes above threshold. See Fig. 13 for an example of this transition in the case of $N = 2$ femtocells.

If instead $c_S = N_T$ and $\gamma_{\mathcal{S}er,S}(\mathbf{a}_k, kT_c) < \gamma_{th}^{\mathcal{S}er}$, the UE starts the handover process to S-BS and the MC evolves to the handover state $< S, 1 >$. As before, the MC crosses deterministically all the handover states $< S, h >$, $h = 2, \ldots, N_H$, and ends up in the cell state where $c_S = \mathrm{C}$ and $c_{S'} = 0$, $\forall S' \neq S$. See Fig. 14 for an example of this transition in the case of $N = 2$ femtocells.

The probability $p_{\mathcal{S}er,S}^{th}(k)$ that the SINR $\gamma_{\mathcal{S}er,S}(\mathbf{a}_k, kT_c)$ is below threshold is computed as in (17) and (18), and is equal to

$$p_{\mathcal{S}er,S}^{th}(k) = \mathrm{P}\left[\gamma_{\mathcal{S}er,S}(\mathbf{a}_k, kT_c) < \gamma_{th}^{\mathcal{S}er}\right] = \frac{\gamma_{th}^{\mathcal{S}er}}{\gamma_{th}^{\mathcal{S}er} + \bar{\gamma}_{\mathcal{S}er,S}(\mathbf{a}_k)} \tag{39}$$

where

$$\bar{\gamma}_{\mathcal{S}er,S}(\mathbf{a}_k) = \frac{\Gamma_{\mathcal{S}er}^{tx}\, g_{\mathcal{S}er}(\mathbf{a}_k)}{\Gamma_S^{tx}\, g_S(\mathbf{a}_k)} \tag{40}$$

is the deterministic part of the SINR $\gamma_{\mathcal{S}er,S}(\mathbf{a}_k, kT_c)$.

Since the received powers from different cells are independent, the transition probabilities among the states of the MC are easily computed from (39) as the product of the probabilities with respect to all cells except the serving one, as can be seen from Fig. 13 and Fig. 14.

As a final comment, we note that the number of states N_{TOT} of the MC described above grows exponentially with the number of femtocells, since

$$N_{TOT} = (N+1)(N_T^N + N_H)\,. \tag{41}$$

However, the complexity of the model can be reduced by considering only transitions among neighboring cells.

REFERENCES

[1] Cisco, "Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2013-2018," *White Paper*, pp. 1–40, Feb. 2014.

[2] J. G. Andrews, "Seven ways that HetNets are a cellular paradigm shift," *IEEE Communications Magazine,* vol. 51, no. 3, pp. 136–144, March 2013.

[3] 3GPP TR 36.839, "Evolved Universal Terrestrial Radio Access (E-UTRA); Mobility enhancements in heterogeneous networks (Release 11)", version 11.0.0, Sept. 2012.

[4] 3GPP TS 36.331, "Protocol Specification; Radio Resource Control," v.10.4.0, Dec. 2011.

[5] D. López-Pérez, I. Guvenc, X. Chu, "Mobility management challenges in 3GPP heterogeneous networks," *IEEE Communications Magazine,* vol. 50, no. 12, pp. 70–78, Dec. 2012.

[6] K. Dimou, M. Wang, Yu Yang, M. Kazmi, A. Larmo, J. Pettersson, W. Muller, Y. Timner, "Handover within 3GPP LTE: Design Principles and Performance," *IEEE Vehicular Technology Conference (VTC Fall)*, pp. 1–5, Sept. 2009.

[7] Q. Liao, S. Stanczak, F. Penna, "A statistical algorithm for multi-objective handover optimization under uncertainties," *IEEE Wireless Communications and Networking Conference (WCNC),* pp. 1552–1557, April 2013.

[8] F. Guidolin, I. Pappalardo, A. Zanella, M. Zorzi, "A Markov-based Framework for Handover Optimization in HetNets," *IEEE Med-Hoc Net,* June 2014.

[9] Q. Liao, F. Penna, S. Stanczak, R. Zhe, P. Fertl, "Context-aware handover optimization for relay-aided vehicular terminals," *IEEE 14th Workshop on Signal Processing Advances in Wireless Communications,* pp. 555–559, June 2013.

[10] M. Peng, D. Liang, Y. Wei, J. Li, H. Chen, "Self-configuration and self-optimization in LTE-advanced heterogeneous networks," *IEEE Communications Magazine*, vol. 51, no. 5, pp. 36–45, May 2013.

[11] D. Xenakis, N. Passas, L. Merakos, C. Verikoukis, "Mobility Management for Femtocells in LTE-Advanced: Key Aspects and Survey of Handover Decision Algorithms," *IEEE Communications Surveys & Tutorials,* vol. 16, no. 1, pp. 64–91, First Quarter 2014.

[12] K. Kitagawa, T. Komine, T. Yamamoto, S. Konishi, "A handover optimization algorithm with mobility robustness for LTE systems," *IEEE 22nd International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC),* pp. 1647–1651, Sept. 2011.

[13] Y. Lee, B. Shin, J. Lim, D. Hong, "Effects of time-to-trigger parameter on handover performance in SON-based LTE systems," *16th Asia-Pacific Conference on Communications (APCC),* pp. 492–496, Nov. 2010.

[14] D. López-Pérez, I. Guvenc, X. Chu, "Theoretical analysis of handover failure and ping-pong rates for heterogeneous networks," *IEEE ICC*, pp. 6774–6779, June 2012.

[15] K. Vasudeva, M. Simsek, I. Guvenc, "Analysis of handover failures in HetNets with layer-3 filtering," *IEEE Wireless Communications and Networking Conference (WCNC)*, pp. 2647–2652, April 2014.

[16] C.H.M. de Lima, M. Bennis, M. Latva-aho, "Modeling and analysis of handover failure probability in small cell networks," *IEEE Conference on Computer Communications Workshops (INFOCOM),* pp. 736–741, May 2014.

[17] X. Lin, R. K. Ganti, P. Fleming, J. G. Andrews, "Towards Understanding the Fundamentals of Mobility in Cellular Networks," *IEEE Transactions on Wireless Communications,* vol. 12, no. 4, April 2013.

[18] K. Kitagawa, T. Komine, T. Yamamoto, S. Konishi, "Performance evaluation of handover in LTE-Advanced systems with pico cell range expansion," *IEEE 23rd International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC),* pp. 1071–1076, Sept. 2012.

[19] B. Jeong, S. Shin, I. Jang, N. W. Sung, H. Yoon, "A Smart Handover Decision Algorithm Using Location Prediction for Hierarchical Macro/Femto-Cell Networks," *IEEE Vehicular Technology Conference (VTC Fall),* pp. 1–5, Sept. 2011

[20] S. Barbera, P. H. Michaelsen, M. Saily, K. Pedersen, "Improved mobility performance in LTE co-channel HetNets through speed differentiated enhancements," *IEEE Globecom,* pp. 426–430, Dec. 2012.

[21] I. Guvenc, "Capacity and Fairness Analysis of Heterogeneous Networks with Range Expansion and Interference Coordination," *IEEE Communications Letters,* vol. 15, no. 10, pp. 1084–1087, Oct. 2011.

[22] N. Zia, S.S. Mwanje, A. Mitschele-Thiel, "A policy based conflict resolution mechanism for MLB and MRO in LTE self-optimizing networks," *IEEE Symposium on Computers and Communication (ISCC),* pp. 1–6, June 2014.

[23] Q. Shen, J. Liu, Z. Huang, X. Gan, Z. Zhang, D. Chen, "Adaptive double thresholds handover mechanism in small cell LTE-A network," *Sixth International Conference on Wireless Communications and Signal Processing (WCSP),* pp. 1–6, Oct. 2014.

[24] A. Lobinger, S. Stefanski, T. Jansen, I. Balan, "Coordinating Handover Parameter Optimization and Load Balancing in LTE Self-Optimizing Networks," *IEEE Vehicular Technology Conference (VTC Spring),* pp.

1–5, May 2011.
[25] F. Guidolin, I. Pappalardo, A. Zanella, M. Zorzi, "Context-Aware Handover in HetNets," *European Conference on Networks and Communications (EuCNC)*, June 2014.
[26] A. Goldsmith, *Wireless Communications,* Cambridge University Press, New York, NJ, 2005.
[27] T. S. Rappaport, *Wireless Communications: Principles & Practice, 2nd Edition,* Prentice Hall, 2002.
[28] N. B. Mandayam, P.-C. Chen, J. M. Holtzman, "Minimum duration outage for cellular systems: a level crossing analysis," *IEEE 46th Vehicular Technology Conference on Mobile Technology for the Human Race,* vol. 2, pp. 879–883, May 1996.
[29] M. Zorzi, "Outage and error events in bursty channels," *IEEE Transactions on Communications,* vol. 46, no. 3, pp. 349-356, March 1998.
[30] R. Tanbourgi, S. Singh, J. G. Andrews, F. K. Jondral, "Analysis of Non-Coherent Joint-Transmission Cooperation in Heterogeneous Cellular Networks," *IEEE ICC,* June 2014.
[31] Q. Ye, B. Rong, Y. Chen, M. Al-Shalash, C. Caramanis, J.G. Andrews, "User Association for Load Balancing in Heterogeneous Cellular Networks," *IEEE Transactions on Wireless Communications*, vol. 12, no. 6, pp. 2706–2716, June 2013.