Modeling and Performance of Uplink Cache-Enabled Massive MIMO Heterogeneous Networks

Anastasios Papazafeiropoulos and Tharmalingam Ratnarajah

***Abstract*—A significant burden on wireless networks is brought by the uploading of user-generated contents to the Internet by means of applications such as the social media. To cope with this mobile data tsunami, we develop a novel multiple-input multiple-output (MIMO) network architecture with randomly located base stations (BSs) a large number of antennas employing cache-enabled *uplink* transmission. In particular, we formulate a scenario, where the users upload their content to their strongest base stations (BSs), which are Poisson point process (PPP) distributed. In addition, the BSs, exploiting the benefits of massive MIMO, upload their contents to the core network by means of a finite-rate backhaul. After proposing the caching policies, where we propose the modified von Mises distribution as the popularity distribution function, we derive the outage probability and the average delivery rate by taking advantage of tools from the deterministic equivalent (DE) and stochastic geometry analyses. Numerical results investigate the realistic performance gains of the proposed heterogeneous cache-enabled uplink on the network in terms of cardinal operating parameters. For example, insights regarding the BSs storage size are exposed. Moreover, the impacts of the key parameters such the file popularity distribution, and the target bitrate are investigated. Specifically, the outage probability decreases if the storage size is increased, while the average delivery rate increases. In addition, the concentration parameter, defining the number of files stored at the intermediate nodes (popularity), affects directly the proposed metrics. Furthermore, a higher target rate results in higher outage because fewer users obey this constraint. Also, we demonstrate that a denser network decreases the outage and increases the delivery rate. Hence, the introduction of caching at the uplink of the system design ameliorates the network performance.**

***Index Terms*—Caching, channel aging, heterogeneous networks, massive MIMO, stochastic geometry**

I. Introduction

A vast majority of new wireless services such as social networks, web-browsing applications, and multimedia streaming has fueled the mobile data traffic with an imminent 500-fold boost over the next 10 years [1]. As a result, mobile operators need to redesign their current networks and delve into more sophisticated techniques to surge system capacity forward and expand coverage in fifth generation (5G) networks [2].

A promising solution towards this direction relies on the deployment of low-power, short-range, and cost-efficient small cell networks (SCNs) or else heterogeneous networks (HetNets). In fact, irregular cellular networks, deployed opportunistically and in hot spots have been researched for a fairly long time now [3]. Specifically, downlink single-input single-output (SISO) HetNets have already presented sufficient progress [4]–[7]. Literally, relevant standardization activities have started in 3GPP release a long time ago [8]. Having assumed that SISO studies are anachronistic, efforts have been devoted to the challenge of modeling multi-antenna HetNets by assuming perfect channel state information (CSI) [9]–[11]. In addition, the practical consideration of imperfect CSI has taken place in several works [12], [13]. Moreover, research has been devoted to the analysis of stochastically geometric uplink models [14], [15].

In a parallel avenue, massive multiple-input multiple-output (MIMO) has emerged as another technology supporting the backbone of 5G networks. Remarkably, massive MIMO point to the increase of spectral and energy efficiencies [16], [17]. The achievement of high cell-throughput along with simple signal processing has been contrived by deploying large-scale antenna arrays at the BS and multi-user (MU) transmission. Starting from the strong assumption of perfect CSI, research has faced realistic impediments such as the presence of pilot contamination [18] and [16], the inevitable hardware impairments [19] and [20], as well as the channel aging [21]–[23]. Especially, by applying the theory of deterministic equivalent (DE) analysis, massive MIMO systems were studied under the presence of channel aging [21]. The key assumption of the DE analysis is that $M \to \infty$ and $K \to \infty$ with a given ratio, where M and K are the numbers of BS antennas and users, respectively. In particular, channel aging refers to the channel variation between the time instance the channel is estimated and the time instance it is used for precoding or detection. The sources of channel aging are mainly the relative movement of the users with the BS antennas, the phase noise, and any processing delays [23]. Despite its significant implications, few works have scrutinized its impact on massive MIMO systems [13], [21]–[23].

The growing trend of user-generated content such as the sharing of real-time events by means of smartphones inflicts a great uploading strain to the wireless networks. Despite the importance of uplink on mobile cellular networks, most efforts

A. Papazafeiropoulos was with the Institute for Digital Communications (IDCOM), University of Edinburgh, Edinburgh, EH9 3JL, U.K. He is now with the Communications and Intelligent Systems Group, University of Hertfordshire, AL10 9AB Hatfield , U. K. and with SnT at the University of Luxembourg, L-1855 Luxembourg. T. Ratnarajah is with IDCOM, University of Edinburgh, Edinburgh. Email: tapapazaf@gmail.com; t.ratnarajah@ed.ac.uk

This work was supported by the U.K. Engineering and Physical Sciences Research Council (EPSRC) under Grant EP/N014073/1 and the UK-India Education and Research Initiative Thematic Partnerships under grant number DST-UKIERI-2016-17-0060.

have been focused on the downlink scenario [24]–[26]. In fact, few attempts have been dedicated to the expanding demands of users' transmission (uploading). Notably, differences appear between the procedures of uploading and downloading. Specifically, an asymmetry regarding the downlink and uplink bandwidths takes place, since the downlink bandwidth can reach $10-1000$ times the corresponding uplink bandwidth. As a consequence, the uploading time is longer, and the throughput is lower. Hence, lower quality of experience is met. Another difference concerns the limited resources of mobile devices such as the battery capacity and the transmit power. Obviously, it is a critical solution to moderate the uplink traffic pressure. An efficient remedy that is brought to the play is the employment of caching in SCNs by exploiting the content popularity appearing as redundancy [27]. In fact, caching the users' content in the edge of the network brings gains regarding the user satisfaction and the traffic load [24]. However, most works address the problem of caching in the downlink direction.

A. Motivation-Central Idea

This paper is motivated by the following observations: 1) Content providers relocate their users' contents from the core network to the intermediate nodes in the network (caching). Hence, a key question to answer is the design and investigation of the converse scenario, where the users move their content to the intermediate nodes to alleviate the upload traffic. 2) We employ a large number of antennas at each BS to take advantage of the benefits of massive MIMO such as the elimination of intra-cell interference. 3) Networks have the tendency to become denser resulting in their irregularity, i.e., it is required to introduce the concept of SCNs. 4) User mobility and its resultant channel aging is a common phenomenon in wireless communications. The main contributions are summarized as follows.

- Contrary to existing works [24], [27], which have studied the downlink caching, we focus on a novel strategy described as *uplink caching*. Hence, instead of having the users requesting contents from their associated BS, the users upload their contents to their strongest BS. More concretely, we formulate the caching problem in the uplink.
- In parallel of considering uplink caching, we take advantage of the gained performance benefits when the massive MIMO technology is employed and HetNet design is encountered. As far as the authors are aware, it is worthwhile to mention that this work is unique regarding the study of the notion of massive MIMO in caching.
- We introduce the modified von Mises distribution instead of a power-law or the Zipf distribution to describe the content popularity distribution, in order to represent the locality and the concentration of the content popularities in a specific region.
- It is the first work introducing channel aging in an architecture including caching. Although we do not show the degradation of the system performance in terms of plots by varying the user mobility due to users' relative movement, the analytical results allow the observation of its dependence and its loss quantification.
- We derive the outage probability and the average delivery rate of an uplink massive MIMO HetNet, where the intermediate nodes are enriched with caching resources. In particular, after having obtained the deterministic signal-to-interference-plus-noise ratio (SINR) by means of the theory of DEs, we achieve to obtain a statistical expression. Notably, the main benefit of the DEs is the provision of deterministic expressions allowing to avoid any Monte Carlo simulations.
- Relied on the numerical results, we elaborate on the impact of various parameters such as the BSs density and the storage size. For example, a high storage size induces improvement of the system, since the outage probability decreases and the average delivery rate increases. For the sake of comparison, we also present the results corresponding to the absence of caching, where applicable.

B. Paper Outline

The remainder of this paper has the following structure. Section II develops the system model of the uplink of a massive MIMO HetNet with channel aging. Section III presents the caching model, while Section IV provides the estimated channel including the effects of pilot contamination and channel aging. Next, Section V presents the uplink transmission under the presence of channel aging and introduction of the caching concept. Section VI provides the main results of this study. Especially, Subsection VI-A includes the derivation and investigation of the outage probability, while Subsection VI-B, provides the presentation of the average delivery rate of this general model. The numerical results are placed in Section VII, and Section VIII concludes the paper.

C. Notation

Vectors and matrices are denoted by boldface lower and upper case symbols. The notations $\mathcal{C}^{M\times 1}$ and $\mathcal{C}^{M\times N}$ refer to complex M-dimensional vectors and $M\times N$ matrices, respectively. The symbols $(\cdot)^{\mathsf{T}}$, $(\cdot)^{\mathsf{H}}$, and $\mathrm{tr}(\cdot)$ express the transpose, Hermitian transpose, and trace operators, respectively. The expectation operator is denoted by $\mathbb{E}[\cdot]$, and the symbol $\triangleq$ declares definition. Moreover, $\mathrm{J}_0(\cdot)$ is the zeroth-order Bessel function of the first kind, and $\Gamma(x,y)$ denotes the Gamma distribution with shape and scale parameters x and y, respectively. Finally, $\mathbf{b}\sim\mathcal{CN}(\mathbf{0},\mathbf{\Sigma})$ represents a circularly symmetric complex Gaussian vector with zero-mean and covariance matrix $\mathbf{\Sigma}$.

II. System Model

This section considers the setup for the uplink of a massive MIMO cellular network consisted of BSs with locations drawn according to an independent PPP Φ_{B} with density λ_{B}, i.e., this formulation corresponds to the generalized and quite interesting design of uplink massive MIMO HetNets. For the sake of better description, a graphical representation of the system layout is shown in Fig 1. Specifically, let the BS of the lth cell having a large number of antennas, denoted by M_l. The users are assumed to be distributed as an independent PPP

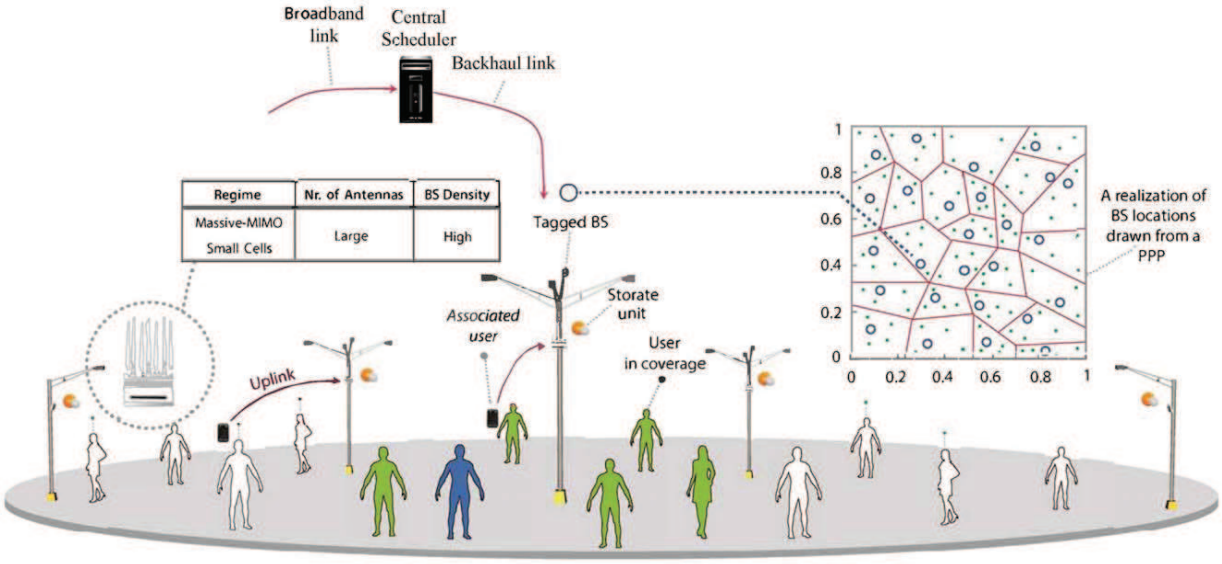

Figure 1: An illustration of the massive MIMO heterogeneous network model with caching. The main figure shows the configuration of BSs, users, storage units, the backhaul, and their interconnection. The rectangular at the top right side illustrates a possible HetNet. The green solid disks and the grey rings represent the users and the BSs, respectively.

with a sufficiently high density λ_{b}. In fact, in any resource block, we assume that the lth BS randomly schedules K_l users according to a distance-based criterion. Actually, the K_l users are connected with the nearest BS constituting its Voronoi cell, while a Voronoi tessellation is structured by the set of all these cells [28][1]. In other words, these users are connected with the strongest BS constituting its Voronoi cell, while a Voronoi tessellation is structured by the set of all these cells[2]. Also, we assume that $M_l \gg K_l$, as stated by the basic principle of massive MIMO technology[3]. Moreover, we consider that each user is equipped with just a single-antenna mobile terminal. Evidently, since the massive MIMO concept is employed, many degrees of freedom are shared across each cell. A central scheduler provides a fixed broadband connection to these BSs by means of wired backhaul links. Obviously, the capacity of the link between the backhaul and each BS is a decreasing function of λ_{B}, since the deployment of more BSs per given area results in less capacity per backhaul link. A capacity expression, obeying to this property, will be introduced in Subsection VI-B.

Further to the network topology, by exploiting Slivnyak's theorem, we are able to conduct the analysis just by focusing on a randomly chosen BS found at the origin [30]. Hereafter, we refer to this BS as the tagged BS. Hence, we assume that at time n, the location of the associated scheduled user is at $x_{llk,n}$, while the location of the kth user found in the jth cell is denoted by $x_{ljk,n}$. Similarly, $\mathbf{h}_{ljk,n} \in \mathbb{C}^{M\times 1}$ is the channel vector from the associated kth user in the cell located at $x_{ljk,n}$ to the tagged BS (located at the origin), while the interference term $\mathbf{h}_{ljk,n} \in \mathbb{C}^{M\times 1}$ is the channel vector corresponding to the link from the kth user of the jth cell located at $x_{ljk,n} \in \mathbb{R}^2$. The locations of the kth scheduled users from all the cells are formed by a non-stationary point process Φ_k, which is not a PPP because of the correlation of their locations with the BS process. The explanation relies on the prohibition of the presence of all other users in Φ_k in the tagged cell [14], [15]. Although this kind of correlations regarding the scheduled users' locations make the exact analysis intractable, we endorse the uplink model, accounting for the pairwise correlations, proposed in [31] and followed in [15]. Moreover, we consider an exclusion ball approximation on the distribution of the scheduled user process Φ_k, being a first-order approximation of the model in [31]. On the top of the ball approximation, we assume that the random variable, expressing the distance $\|x_{llk,n}\|^{-\alpha}$ from the scheduled user to its tagged BS at the origin, is assumed to be a Rayleigh variable with a mean of $0.5\sqrt{1/\lambda_{\mathrm{B}}}$ [14]. We denote $R_e = \sqrt{1/(\pi\lambda_{\mathrm{B}})}$ the radius of the ball, in order to let the size of the surface of the exclusion ball equal the average cell size, which is $1/\lambda_{\mathrm{B}}$ [32]. Furthermore, the scheduled user process Φ_k, describing the locations of the other scheduled users, is formed as an inhomogeneous PPP of density λ_k where the users are found outside an exclusion ball having as center the tagged BS. Especially, the locations of the kth users of the other cells, belonging to Φ_k, are modeled by means of an inhomogeneous PPP with a density function of

$$\lambda_k(r) = \lambda_{\mathrm{B}}\left(1 - \mathrm{e}^{-\lambda_{\mathrm{B}}\pi r^2}\right), \tag{1}$$

where $r = \|x_{ljk,n}\|$.

[1]In this work, we assume only a non-line-of-sight (NLoS) transmission, while the consideration of an LoS component is left for future work. Its introduction in the analysis could be made by means of a multi-slope path-loss model [29].

[2]The users are assumed to be distributed as an independent point process, but each cell is large enough (the density of the users' PPP is sufficiently large) to shelter K_l users. The users in each cell are independently and uniformly distributed.

[3]Although, in practice the number of BS antennas M_l, and the number of associated users K_l differ across cells, henceforth, we assume $M_l = M$ and $K_l = K$ for the sake of simplicity.

Both the uplink training and data transmission phases necessitate the introduction of fractional power control in our analysis. Thus, the k user in the the lth cell transmits with power

$$P_{lk,n} = P_{\mathrm{t}} \beta_{llk,n}^{-\epsilon}, \tag{2}$$

where $\epsilon \in [0,1]$ expresses the fraction of the compensation of the path-loss given by $\beta_{llk,n}^{-\epsilon}$, while P_{t} is the open loop transmit power assuming no power control.

As far as the channel model is concerned, let the point-to-point channels be characterized by independent and identically distributed (i.i.d.) Rayleigh block fading with unit mean, while we assume a block fading model, where the channel is assumed constant during one block, but varies independently from block to block. Note that although the assumption of both line and non-line of sight signals appear in small cells, we consider only Rayleigh fading for the sake of simplicity. Relaxation of the Rayleigh fading assumption as well as the introduction and study of other fading models can be done with techniques found in [33], and is left for future work. Hence, in the proposed model, the channel vector $\mathbf{h}_{ljk,n}$ from user k in the jth cell at the n-th time slot is modelled as

$$\mathbf{h}_{ljk,n} = \beta_{ljk}^{1/2} \mathbf{w}_{ljk,n}, \tag{3}$$

where β_{ljk} is the large-scale path-loss and $\mathbf{w}_{ljk,n} \in \mathbb{C}^{M \times 1}$ is an uncorrelated fast fading Gaussian channel vector with elements having zero mean and unit variance, i.e., $\mathbf{w}_{ljk} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_M)$. Note that the incorporation of spatial correlation due to lack of limited antenna spacing, and different antenna patterns is left for future work. The path-loss at the tagged cell is described by

$$\beta_{ljk} = C r_{ljk}^{-\alpha}, \tag{4}$$

where $\alpha > 2$ is the path-loss exponent, and C expresses a constant determined by the carrier frequency and the reference distance. Given that we have assumed an NLoS component, the distance-based criterion is translated to path-loss based, i.e., the minimum distance corresponds to the minimum path loss signal. For the sake of exposition, we have assumed a simplistic single-slope path-loss model, in order not to distract the reader from the main contributions. The application of a more complex model, such as the multi-slope path-loss model presented in [34], is outside of the scope of this paper and left to future work.

The transmission scheme includes an uplink channel estimation phase, and continues with an uplink data transmission phase, allowing the derivation of the outage probability and the average delivery rate, in order to shed light on their behavior. However, we need first to introduce the caching model.

III. Caching Model

There are definitely certain cases that BSs need to cache the files of users in the uplink and backhaul for further cost savings, latency reduction, etc. For example, imagine a crowded scenario where many users are willing to upload their video recordings to the Internet/network. If uploaded files could be proactively cached at the BSs, that could be beneficial to the network if nearby users have later the interest to download/watch those uploaded videos. Proactive caching at the uplink could alleviate the upload traffic. This can be also a criterion to select which file is of high interest. Specifically, assuming that the nodes/users upload files via uplink, there is a chance that the user who is uploading the file will have many downloads (i.e., Justin Bieber sharing a video, most likely will be viewed by his followers). Therefore, the file of interest could be inferred based on proactive prediction methods of how the uploading device is "influential", relying on machine learning, social networks, etc. Moreover, although there is no not too much study about the role of caching in the uplink, uplink caching will be very beneficial when the Internet of Things (IoT) devices will be introduced on cellular networks [35, Fig. 37].

Let us consider that the network has a *content catalog* of F contents represented by the set of $\mathcal{F} = \{f_1, ..., f_F\}$. User k in jth cell at the n-th time slot (located at $x_{ljk,n}$) demands a content from a sub-catalog $\mathcal{F}_l \subseteq \mathcal{F}$ according to a content popularity distribution f_{pop}. In particular, the BS at the tagged cell has a content popularity distribution f_{pop} and is modelled by a *modified* von Mises distribution [36], which is a symmetric circular distribution defined as

$$f_{\mathrm{pop}}(f, \mu_j, \gamma_j) = \frac{e^{\gamma_j \cos\left(\left(2\pi(f-\mu_j)/F\right)-\pi\right)}}{2\pi J_0\left(\gamma_j\right)}, \tag{5}$$

where f is a point in the support such as $0 \leq f \leq F$, the parameter μ_j is a measure of location such as $0 \leq \mu_j \leq F$, the parameter γ_j is a measure of concentration with $0 \leq \gamma_j \leq +\infty$, and the function $J_0\left(\gamma_j\right)$ is the zeroth-order Bessel function of the first kind. The distribution becomes uniform when $\gamma_j = 0$ and highly concentrated on the point μ_j when $\gamma_j \to \infty$. The parameters μ_j and $1/\gamma_j$ are analogous to the mean and variance in Gaussian distribution. In fact, when $\gamma_j \to \infty$, we obtain the Dirac delta function. The intuition behind such a distribution and modelling is due to the observation that the content catalog is finite $[0, F]$ and the content popularities might be concentrated on specific region, where parameters μ_j and $1/\gamma_j$ are used for its description. Suppose that the j-th BS has a storage capacity of S_j nats with $S_j \leq F$ (1 bit = $\ln(2) = 0.693$ nats), and caches files according to the policy provided below. Henceforth, all the parameters concerning caching are the same across all cells, e.g., we assume that $\mu_j = \mu$ and $\gamma_j = \gamma$. The length of each file in the catalog is L nats, while T expresses its bitrate in nats/s/Hz. Note that the uplink rate of each user has to be equal or higher than the file bitrate T, in order to avoid any interruption during its experience.

We assume that we store the most popular files from the catalog in advance offline. Storing most popular files requires perfect knowledge of the content popularity, which might not be possible to be constructed locally. In order to make the things local/geographical, an interesting caching policy is to store the Sth closest different files mentioned above.

If the file is of high interest for the BS (if it will be a popular file in the future and is not cached yet), then the BS should cache it, i.e., uplink transmission incurs. If the file is of high interest for the BS (if it will be a popular file in the future) and is already cached, then the BS should do nothing (cache miss).

In other words, in such case, cache hit means that the user is in coverage and the file is not included in the BS. Hence, uploading to the BS is meaningful; otherwise, if the user is not in coverage or the file is included at the BS, the file will not be uploaded or it will be uploaded to the core network from the BS.

Notably, the uplink caching process is dynamic, since the users are likely to have a popular content at any time, which should be uploaded for the better performance of the network. Hence, the proposed model is constantly vital. However, even when all the users have uploaded their contents, they are able to exploit the model and focus on downloading a content that already has been uploaded. The latter scenario is very rare because it is out of chance that at some point all the users will have uploaded their contents. At any time, there will be at least one user that will have some popular content to upload.

IV. Channel Estimation

Let us denote the channel coherence time is T_{c}. We assume that the same time-frequency resources are shared by the users across all cells. Aiming to the characterization of realistic systems, we account for imperfect CSIT due to pilot contamination and channel aging. Let τ denote the length of the training period. Obeying to time-division duplex (TDD) design, during the uplink training phase, having duration τ symbols, the tagged BS obtains the estimated channel. The uplink data transmission phase consists of $T_{\mathrm{c}}-\tau$ symbols.

Having in mind that the signal from each user is attenuated with distance because of the path-loss, we present the pilot contamination occurred due to the re-use of the pilot sequences during the training phase.

A. Pilot Contamination

According to TDD, estimation of the local CSI takes place during the uplink training phase, where the same band of frequencies is shared across all cells. Moreover, the kth user in each cell is assigned with the same pilot sequence. As a result, pilot contamination occurs and the degradation of the system performance is inevitable. Let the superscript tr describe the training stage. Furthermore, the scheduled user processes with different pilots, i.e., Φ_k, $\Phi_{k'}$, are assumed to be independent. The tagged BS receives a noisy observation of the channel vector from the associated sheduled user at time instance n. The average power of each transmitted pilot symbol from the sheduled user is $P_{lk,n}$. Hence, the associated BS observes the channel $\mathbf{h}_{llk,n}$ as

$$\mathbf{y}^{\mathrm{tr}}_{llk,n}=\underbrace{\sqrt{P_{lk,n}}\mathbf{h}_{llk,n}}_{\substack{\text{Desired}\\ \text{signal}}}+\underbrace{\sum_{j\neq l}\sqrt{P_{jk,n}}\mathbf{h}_{ljk,n}}_{\substack{\text{Interference}\\ \text{part}}}+\underbrace{\mathbf{N}^{\mathrm{tr}}_{llk,n}\boldsymbol{\psi}^{\mathrm{H}}_{k}}_{\text{noise}}, \quad (6)$$

where the vector $\boldsymbol{\psi}_k \in \mathbb{C}^{\tau\times 1}$ denotes the training sequence of the kth user with $\boldsymbol{\psi}_k\boldsymbol{\psi}^{\mathrm{H}}_k$=1, and $\mathbf{N}^{\mathrm{tr}}_{llk,n} \in \mathbb{C}^{M\times\tau}$ is the spatially white additive Gaussian noise matrix with i.i.d. entries distributed as $\mathcal{CN}\left(0,\frac{\sigma^2}{K}\right)$. Note that the channel vectors $\mathbf{h}_{ljk,n}$, being independent across cells and user distances, are Gaussian distibuted as $\mathcal{CN}\left(\mathbf{0},\mathbf{I}_M\right)$.

The tagged BS estimates $\mathbf{h}_{llk,n}$ by applying minimum mean square error (MMSE) estimation to (6), and by assuming that the tagged BS knows perfectly the large-scale path-losses β_{ljk} for $j\neq l$. Thus, the estimated channel is

$$\begin{aligned}\hat{\mathbf{h}}_{llk,n}\Big|_{\|r_{ljk,n}\|} &= \mathrm{E}\Big[\mathbf{h}_{llk,n}\mathbf{y}^{\mathrm{tr,H}}_{llk,n}\Big]\mathrm{E}^{-1}\Big[\mathbf{y}^{\mathrm{tr}}_{llk,n}\mathbf{y}^{\mathrm{tr,H}}_{llk,n}\Big]\mathbf{y}^{\mathrm{tr}}_{llk,n}\\ &= \frac{\sqrt{P_{lk,n}}\beta_{llk}}{\sum_j P_{jk,n}\beta_{ljk}+\frac{\sigma^2}{K}}\mathbf{y}^{\mathrm{tr}}_{llk,n}, \quad (7)\end{aligned}$$

and it is distributed as $\hat{\mathbf{h}}_{llk,n}\sim\mathcal{CN}\left(\mathbf{0},\sigma^2_{\hat{\mathbf{h}}_k}\mathbf{I}_M\right)$ with variance given as

$$\sigma^2_{\hat{\mathbf{h}}_{llk}}=\frac{P_{lk,n}\beta^2_{llk}}{\sum_j P_{jk,n}\beta_{ljk}+\frac{\sigma^2}{K}}. \quad (8)$$

Based on the orthogonality principle of the MMSE estimation, the uncorrelated estimation error vector at time instance n is $\hat{\mathbf{e}}_{llk,n}=\mathbf{h}_{llk,n}-\hat{\mathbf{h}}_{llk,n}$, being distributed as $\hat{\mathbf{e}}_{llk,n}\sim\mathcal{CN}\left(\mathbf{0},\sigma^2_{\hat{\mathbf{e}}_{llk}}\mathbf{I}_M\right)$ with

$$\sigma^2_{\hat{\mathbf{e}}_k}=\beta_{llk}\left(1-\frac{P_{lk,n}\beta_{llk}}{\sum_j P_{jk,n}\beta_{ljk}+\frac{\sigma^2}{K}}\right). \quad (9)$$

B. Channel Aging

The relative movement of the kth associated user with a comparison to the tagged BS antennas results in the variation of the channel. Hence, this source of imperfection contributes further to the need for estimation of the channel. Mathematically, we are able to relate the current sample of the channel with its past samples by means of an autoregressive model of order s [37]. Herein, for the sake of computational complexity and tractability, we choose an autoregressive model of order 1, which is a common approach in the literature [38]. Thus, the current channel at the tagged BS is modeled as

$$\mathbf{h}_{llk,n}=\delta\mathbf{h}_{llk,n-1}+\mathbf{e}_{llk,n}, \quad (10)$$

where $\mathbf{h}_{llk,n-1}$ is the channel in the previous symbol duration, and $\mathbf{e}_{llk,n}\in\mathbb{C}^N$, modelled as a stationary Gaussian random process with i.i.d. entries and distribution $\mathcal{CN}(\mathbf{0},(1-\delta^2)\mathbf{I}_M$, is the uncorrelated channel error because of the channel variation [38]. Regarding δ, it is related to the second-order statistics. Specifically, an appropriate measure for modeling the variation of the channel is its second-order statistics, which can be described by means of the autocorrelation function of the channel. For this role, a widely accepted model is the Jakes model due to its generality and simplicity [37]. The Jakes model describes a propagation medium with two-dimensional isotropic scattering and a monopole antenna at the receiver [39]. In such case, the normalized discrete-time autocorrelation function of the fading channel is expressed by

$$r\left(s\right)=J_0(2\pi f_D T_s|s|), \quad (11)$$

where f_D and T_s are the maximum Doppler shift and the channel sampling period. Especially, the maximum Doppler shift f_D can be expressed by means of the relative velocity of

the scheduled user v, i.e., $f_D = \frac{vf_c}{c}$, where $c = 3 \times 10^8$ m/s is the speed of light and f_c is the carrier frequency. Also, s denotes the delay. Increasing the argument of the Bessel function results in a decrease of the magnitude to zero but with some ripples in the meanwhile. We set $\delta = r[1]$, i.e., we consider a single symbol delay. To this end, we assume that the BS has perfect knowledge of δ.

Remarkably, following the procedure in [21], we are able to write both pilot contamination and time-variation of the channel as a combination. More concretely, the channel at time slot n can be written as

$$\begin{aligned}\mathbf{h}_{llk,n} &= \delta \mathbf{h}_{llk,n-1} + \mathbf{e}_{llk,n} \\ &= \delta \hat{\mathbf{h}}_{llk,n-1} + \tilde{\mathbf{e}}_{llk,n}, \end{aligned} \tag{12}$$

where $\hat{\mathbf{h}}_{llk,n-1}$ and $\tilde{\mathbf{e}}_{llk,n} = \delta\tilde{\mathbf{h}}_{llk,n-1} + \mathbf{e}_{llk,n} \sim \mathcal{CN}\left(\mathbf{0}, \sigma^2_{\tilde{\mathbf{e}}_k}\mathbf{I}_M\right)$ with $\sigma^2_{\tilde{\mathbf{e}}_k} = \left(1 - \delta^2\sigma^2_{\hat{\mathbf{h}}_k}\right)$ are mutually independent. Hence, the estimated channel of the k scheduled user at time n is provided by $\hat{\mathbf{h}}_{llk,n} = \delta\hat{\mathbf{h}}_{llk,n-1}$. Note that in the special case, where $\delta = 1$, we obtain a static environment with no user mobility.

V. Uplink Transmission

In general, the physical representation of a link defines the probability distribution function (PDF) of this link. Specifically, we face different distributions depending if we model the desired or the interference part of the received signal. Another example, affecting the PDF, concerns the choices between multi-antenna and single-antenna BS architecture, and between single or multi-user transmission. Notable, herein, we employ the general setting of a large number of antennas deployed by the tagged BS serving multiple users simultaneously. The first step towards the statistical characterization of the powers of the received signal's parts is to model the uplink transmission.

Thus, accounting for a quasi-static block fading model with frequency-flat fading channels varying for symbol to symbol, the received signal from the associated scheduled user at $x_{llk,n}$ to the tagged BS during the nth time-slot after applying a general decoder $\mathbf{q}_{llk,n}$ can be expressed as

$$\begin{aligned}y_{llk,n} = \mathbf{q}^{\mathsf{H}}_{llk,n}\mathbf{h}_{llk,n}s_{lk,n} + \sum_{(j,k')\neq(l,k)} \mathbf{q}^{\mathsf{H}}_{llk,n}\mathbf{h}_{ljk',n}s_{jk',n} \\ + \mathbf{q}^{\mathsf{H}}_{llk,n}\mathbf{n}_{llk}, \end{aligned} \tag{13}$$

where $s_{lk,n}$ is the uplink data symbol of the kth scheduled user with $\mathbb{E}\left[|s_{lk,n}|^2\right] = P_{lk,n}$. The channel vector $\mathbf{h}_{llk,n} \in \mathbb{C}^{M\times 1}$ denotes the desired channel vector between the tagged BS and the associated kth scheduled user located at $r_{llk,n} \in \mathbb{R}^2$ at time-instace n. Similarly, $\mathbf{h}_{ljk',n} \in \mathbb{C}^{M\times 1}$ expresses the interference channel vector from the other users found at $r_{ljk',n} \in \mathbb{R}^2$ far from the typical BS at time-instace n. Also, $\mathbf{n}_{llk} \in \mathbb{C}^{M\times 1} \sim \mathcal{CN}\left(0, \sigma^2\mathbf{I}_M\right)$ is the Gaussian thermal noise vector in the uplink data transmission.

Taking into account for the realistic case, where imperfect CSI due to pilot contamination and time-variation of the channel (see (12)), is considered, the received signal by the tagged BS can be written as

$$\begin{aligned}y_{llk,n} = \delta\mathbf{q}^{\mathsf{H}}_{llk,n}\hat{\mathbf{h}}_{llk,n-1}s_{lk,n} + \mathbf{q}^{\mathsf{H}}_{llk,n}\tilde{\mathbf{e}}_{llk,n}\mathbf{s}_{k,n} \\ + \sum_{(j,k')\neq(l,k)} \mathbf{q}^{\mathsf{H}}_{llk,n}\mathbf{h}_{ljk',n}s_{jk',n} + \mathbf{q}^{\mathsf{H}}_{llk,n}\mathbf{n}_k, \end{aligned} \tag{14}$$

where we have replaced the current channel by means of (12) with its estimated version[4]. In (14), the first term expresses the desired signal received by the tagged BS. The second term describes the estimation error effect. Furthermore, the third term represents the other users interference, while the last term denotes the post-processed noise.

The achievable uplink SINR from the k scheduled user to the tagged BS, denoted by SINR_k, is shown in (16), where we have treated the unknown terms at the tagged BS as uncorrelated additive noise. The encoding of the message takes place over many realizations of certain sources of randomness in the model. Specifically, the expectation operators are taken over the channel estimation error, the small-scale fading in the interference links, and the thermal noise. Notably, the resultant SINR, provided by (17), is a random variable because of the randomness accompanying the large-scale path-losses. Thus, we result in the uplink of a multi-user large MIMO HetNet, where the SINR expression will be investigated by using tools from stochastic geometry and large random matrix theory.

Herein, we present the derivation of an approximation of the SINR, when maximal-ratio combining (MRC) receivers are employed. As the number of BS antennas grows large, the approximation becomes tighter according to the theory of DEs [40]. Starting from the DE SINR distribution, we derive below the outage probability and the average delivery rate.

Let the tagged BS apply the decoder $\mathbf{q}_{llk,n}$ to the received signal, being a scaled version of the channel estimate $\delta\hat{\mathbf{h}}_{llk,n-1}$. The mathematical expression of the MRC decoder is

$$\mathbf{q}_{llk,n} = \frac{\sum_j P_{jk,n}\beta_{ljk} + \frac{\sigma^2}{K}}{\sqrt{P_{lk,n}}\beta_{llk}}\delta\hat{\mathbf{h}}_{llk,n-1}, \tag{15}$$

where the scaling of the decoder is applied for the sake of simplicity, but it will not affect the SINR distribution. Also, note that the decoder depends on the estimated channel, obtained during the training phase. Hereafter, we omit the time index from the expressions, while note that the DE expressions are calculated over the channel distributions, i.e., they are conditioned on the BSs positions.

Let $\overline{\mathrm{SINR}}_{llk}$ be the deterministic SINR, obtained such that $\mathrm{SINR}_{llk} - \overline{\mathrm{SINR}}_{llk} \xrightarrow[M\to\infty]{\text{a.s.}} 0$[5].

Proposition 1. *The uplink achievable DE SINR with MRC decoding under the presence of pilot contamination and channel aging is given by* (17) *with* $W_{jk} = \beta^{-\epsilon}_{jjk}\beta_{ljk}$.

Proof: See Appendix B. ∎

Notably, the terms W_{jk} correspond to the interference terms from other cells.

[4]Note that the replacement concerns only the current desired channel because the interference part is not of direct interest and can be seen as additive noise.

[5]The notation $\xrightarrow[M\to\infty]{\text{a.s.}}$ denotes almost sure convergence, while the definition of the term "deterministic equivalent" is given by [40, Def. 6.1].

$$\mathrm{SINR}_{llk,n}=\frac{\delta^2 P_{lk,n}|\mathbf{q}_{llk,n}^{\mathsf{H}}\hat{\mathbf{h}}_{llk,n}|^2}{P_{lk,n}\mathbb{E}\left[|\mathbf{q}_{llk,n}^{\mathsf{H}}\tilde{\mathbf{e}}_{llk,n}|^2\right]+\sum_{(j,k')\neq(l,k)}P_{jk',n}\mathbb{E}\left[|\mathbf{q}_{llk,n}^{\mathsf{H}}\mathbf{h}_{ljk',n}|^2\right]+\|\mathbf{q}_{llk,n}\|^2\sigma^2}. \tag{16}$$

$$\overline{\mathrm{SINR}}_{llk}=\frac{P_{\mathrm{t}}\delta^2\beta_{llk}^{2(1-\epsilon)}}{P_{\mathrm{t}}\beta_{llk}^{(1-\epsilon)}\left(\sum_{j\neq l}W_{jk}+\frac{\sigma^2}{P_{\mathrm{t}}K}\right)-P_{\mathrm{t}}\delta^2\beta_{llk}^{2(1-\epsilon)}+\sum\limits_{(j,k')\neq(l,k)}P_{\mathrm{t}}\left(\left(\frac{\sigma^2}{P_{\mathrm{t}}K}+W_{jk}\right)W_{jk'}+W_{jk}^2\right)+\sum\limits_{j}W_{jk}+\frac{\sigma^2}{P_{\mathrm{t}}K}}. \tag{17}$$

VI. Performance Analysis

The proposed realistic system depends on several practical factors, e.g., the pilot contamination and the channel aging as well as caching parameters such as the storage size. As already known, the quality-of-experience constraints specify that the uplink rate of the scheduled user should be equal or higher than the file bitrate T so that the user does not observe any interruption during its experience. In addition, another impediment is not been taken into account in most cases. It concerns the rate of backhaul, which becomes quite important when the cache misses.

The quantification and the assessment of the system necessitate the definition of certain metrics, namely the outage probability and the average delivery rate.

A. Outage probability

In this section, we present the uplink outage probability of the associated user in a large antenna MU HetNet with imperfect CSIT due to pilot contamination and channel aging, while caching is employed. The technical derivation is given in Appendix C. As a performance metric, the outage probability is given as the complementary of the success (coverage) probability, expressing the joint probabilities of the uplink rate exceeding the file bitrate T and the received file missing from the local cache. It is worthwhile to mention that only the BSs connect with the backhaul. In other words, the associated user is not able to upload its content if the BS already has it. Actually, there is no reason to do it, since the content is stored at the BS and it is BS's task to upload it through its wired link. Hence, we have

$$\mathbb{P}_{\mathrm{out}}\triangleq 1-\tilde{\mathbb{P}}\left(\overline{\mathrm{SINR}}_{llk}>\tilde{T},f\notin\Delta_l\right), \tag{18}$$

where $\tilde{T}=e^T-1$, f is the received file by the typical BS, and Δ_l is the local cache of the served BS at the lth cell. Differently to [27], this definition follows another line of reasoning. In particular, if the requested file is not in the cache of the served BS, and if the uplink rate is higher than the file bitrate T, then, the user uploads its content and does not observe any interruption during its communication. Hence, we expect the outage probability to be close to zero. Formally, the outage probability is given by the following theorem.

Theorem 1. *The approximated uplink outage probability $\mathbb{P}_{\mathrm{out}}$ in a large MU-MIMO HetNet with caching attributes, accounting for imperfect CSIT due to pilot contamination and channel aging, is given by*

$$\mathbb{P}_{\mathrm{out}}\approx 1-\left(1-\int_0^{\frac{S}{L}}f_{\mathrm{pop}}(f,\mu,\gamma)\mathrm{d}f\right)\tilde{\mathbb{P}}\left(\overline{\mathrm{SINR}}>\tilde{T}\right) \tag{19}$$

with $f_{\mathrm{pop}}(f,\mu,\gamma)$ given by (5) *and the coverage probability $\tilde{\mathbb{P}}\left(\overline{\mathrm{SINR}}>\tilde{T}\right)$ given by* (21) *with $t=\pi\lambda_B x^2$. The variable N represents the number of terms used in the calculation, $\eta=N(N!)^{-\frac{1}{N}}$, while $D_{\sigma^2}=\frac{\sigma^2}{KP_{\mathrm{t}}C^{1-\epsilon}(\lambda_{\mathrm{b}}\pi)^{\frac{\alpha(1-\epsilon)}{2}}}$, $D_1=\frac{2\Gamma^{\alpha}\left(\frac{\epsilon}{2}+1\right)}{\alpha-2}+D_{\sigma^2}$, $D_2=(K-1)D_1$, $D_3=(K-1)\frac{\Gamma^{\alpha}(\epsilon+1)}{\alpha-1}$, $C_4=D_1+2C^{1-\epsilon}(\pi\lambda_{\mathrm{b}})^{\frac{1-\alpha\epsilon}{2}}\Gamma^{\alpha}\left(\frac{\epsilon}{2}+1\right)$, $D_5(t)=\int_0^{\infty}\frac{\mathrm{e}^{-u}\mathrm{d}u}{1+D_2t^{2\alpha(1-\epsilon)}u^{-\frac{\alpha}{2}(1-\epsilon)}}$, and $D_6=-\frac{1}{P_{\mathrm{t}}}$.*

Proof: See Appendix C. ∎

B. Average Delivery Rate

This section presents the derivation of the average delivery rate, defined as

$$R\triangleq\begin{cases}T, & \text{if } \psi\ln\left(1+\overline{\mathrm{SINR}}_{llk}\right)>T \text{ and } f\notin\Delta_{b_0}\\ C(\lambda_{\mathrm{B}}), & \text{if } \psi\ln\left(1+\overline{\mathrm{SINR}}_{llk}\right)>T \text{ and } f\in\Delta_{b_0}\\ 0, & \text{otherwise}\end{cases} \tag{20}$$

where $C(\lambda_{\mathrm{B}})=\frac{C_1}{\lambda_{\mathrm{B}}}+C_2$ with C_1 and C_2 being arbitrary coefficients under the constraint the ceiling of the delivery rate is $C(\lambda_{\mathrm{B}})$ with $C(\lambda_{\mathrm{B}})<T$. $C(\lambda_{\mathrm{B}})$ denotes the backhaul capacity being available to the intermediate nodes. Also, $\psi=\frac{T_{\mathrm{c}}-\tau}{T_{\mathrm{c}}}$ is the fraction of time expressing the training overhead which occurs during the estimation channel. However, (20) refers to the opposite direction (uplink) and then it includes an interesting and insightful explanation, especially, because the conditions are different. In the case that the uplink rate is higher than the target file rate (bitrate) T and the file is not found in the BSs, the user uploads at full rate T. On the contrary, if the rate is greater than T and the file is already to the local cache, the associated user does not upload its content, but the tagged BS does. The latter constraint relies on the assumption that a high-speed backhaul is not cost-efficient in dense networks.

Theorem 2. *The approximated uplink average delivery rate of the typical BS in a large MU-MIMO HetNet with caching attributes, accounting for imperfect CSIT due to pilot contamination and channel aging, is given by*

$$R=\left(T-(C(\lambda_{\mathrm{B}})-T)\int_0^{\frac{S}{L}}f_{\mathrm{pop}}(f,\mu,\gamma)\mathrm{d}f\right)\tilde{\mathbb{P}}\left(\overline{\mathrm{SINR}}>T\right) \tag{22}$$

$$\tilde{\mathbb{P}}\left(\overline{\mathrm{SINR}} > \tilde{T}\right) \approx \sum_{n=1}^{N} \binom{N}{n} (-1)^{n+1} \int_0^{\infty} \mathrm{e}^{-t - n\eta\tilde{T}\delta^{-2}\left(D_1 t^{\alpha(1-\epsilon)} + D_3 t^{2\alpha(1-\epsilon)} + D_4 t^{2\alpha(1-\epsilon)} + D_6\right)} \left(1 - D_2 t^{2\alpha(1-\epsilon)} D_5\left(t\right)\right)^{K-1} \mathrm{d}t. \tag{21}$$

where $f_{\mathrm{pop}}(f, \mu, \gamma)$ is given by (5)*, and the coverage probability is given by* (21) *with $t = \pi\lambda_B x^2$ if we substitute $\tilde{T}$ with T.*

Proof: See Appendix D. ∎

VII. Numerical Results

In this section, we illustrate the behavior of the analytical expressions concerning the outage probability $\mathbb{P}_{\mathrm{out}}$ and the average delivery rate R, which are provided by means of (19) and (22)[6]. In fact, we investigate the impact of various design parameters such as the BS density λ_{B}, the storage size of BSs S in nats, and the target bit-rate T in nats/sec/Hz. Also, the analytical expressions are verified by Monte Carlo simulations. The simulated curves were obtained by averaging the corresponding expressions over 10^3 random instances. Actually, the simulated results of the outage probability $\mathbb{P}_{\mathrm{out}}$ and the average delivery user rate R are depicted along with the proposed analytical expressions. Specifically, the bullets correspond to the simulation results, while the "solid" lines represent the proposed analytical results by varying their parameters. The discrimination between "solid" and "dot" lines, where applicable, designates the results with "caching" and "no caching", respectively. The "no caching" scenario is obtained by assuming that the content popularity distribution coincides with the Dirac delta function, i.e., $\gamma_j \to \infty$.

The simulations are conducted by following a specific procedure. Specifically, we choose a sufficiently large area of $5\ \mathrm{km} \times 5\ \mathrm{km}$, where the locations of the BSs are simulated as a realization of a PPP with given density $\lambda_{\mathrm{B}} = 0.2\ \mathrm{m}^{-2}$. Next, the users' PPP density is considered to be $\lambda_{\mathrm{K}} = 60\lambda_{\mathrm{B}}$[7]. The association relies on the minimum path-loss (distance-based) rule, while K users from each cell are randomly scheduled. Hence, we select the strongest user to the tagged BS, found at the origin, as the associated scheduled user at $x_{llk,n}$. It is worthwhile to mention that the users could employ other schemes to upload their contents to the BSs. For example, they could select the serving BSs rather than the closest BSs. The relevant comparison with other approaches regarding the selection of the appropriate BSs is interesting and is left for future work. Furthermore, the setup includes BSs of $M = 25$ number of antennas, while we pick $K = 5$ users per BS. The system under study, embodying a such number of BS antennas, is considered to describe a massive MIMO model, since the simulations coincide with the DEs. In other words, the DEs are tight approximations even for this number of antennas. Hence, such a number of BS antennas can represent a massive MIMO model. However, this is not a new observation. According to the literature, Similar observations have been made in the literature even for an 8×8 system [40]–[43]. The average uplink transmit power for both training and transmission phases is $P_{lk,n} = 2$ dBW, and the bandwidth allocated for each user is

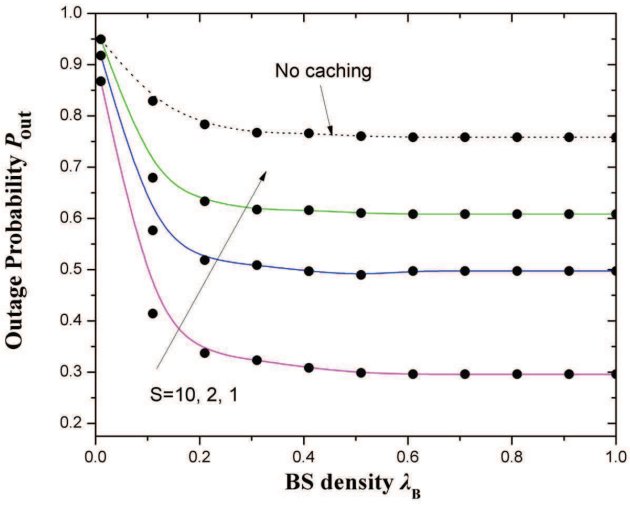

Figure 2: Outage probability versus the BS density λ_{B} for varying storage size S. Solid lines and bullets correspond to the theoretical and simulated results, respectively, while the dotted line refers to the "No caching" scenario.

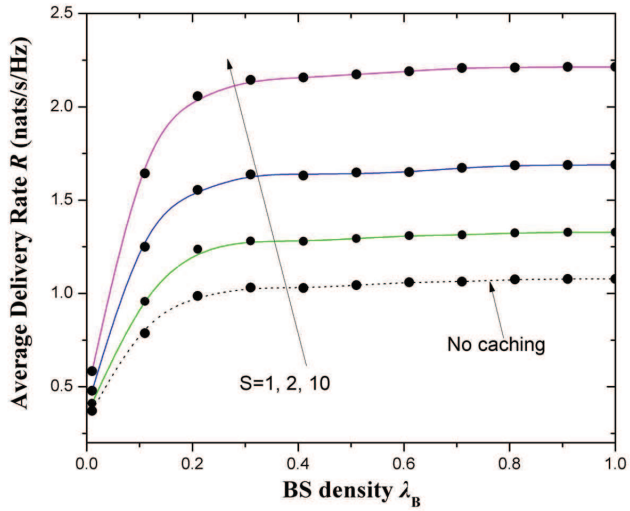

Figure 3: Average delivery rate versus the BS density λ_{B} for varying storage size S. Solid lines and bullets correspond to the theoretical and simulated results, respectively, while the dotted line refers to the "No caching" scenario.

[6]Remarkably, there is no known result in the literature studying caching in the uplink of a HetNet employing a large number of antennas (massive MIMO). In addition, there is no known reference investigating channel aging in the case of cached-enabled BSs.

[7]Although the analytical expressions rely on the assumption of an infinite plane, the simulation takes place over a finite window.

20 MHz. Also, regarding the rest parameters, we set $L = 1$ nats, $\alpha = 3$, $P_{\rm t} = 1$, $\epsilon = 0.7$, $C_1 = 0.005$, $C_2 = 0$, $\delta = 0.8$, $\mu = 0$, $\gamma = 0.5$ unless otherwise stated. Due to limited space, in this work, we do not focus on channel aging, studied in other works such as [13], but the cynosure is the impact of caching in the uplink.

A. Impact of BS Density

In Fig. 2, we illustrate the behavior of the outage probability $\mathbb{P}_{\rm out}$ with respect to the BS density $\lambda_{\rm B}$ for different values of the storage size S. We observe a decrement of the outage probability as the BS density increases. In other words, a denser HetNet provides better coverage. At the same time, an increase of the storage size of the intermediate nodes brings a decrease in an outage, since the users do not have to upload their content to the core network because the BSs have plenty of space to save the receiver information.

Regarding the average delivery rate R, it increases with the BS density $\lambda_{\rm B}$ as can be seen in Fig. 3. However, it saturates soon due to the increasing intra-cell interference. Moreover, higher storage size contributes to the increase of the rate because of the traffic load towards the backhaul is alleviated.

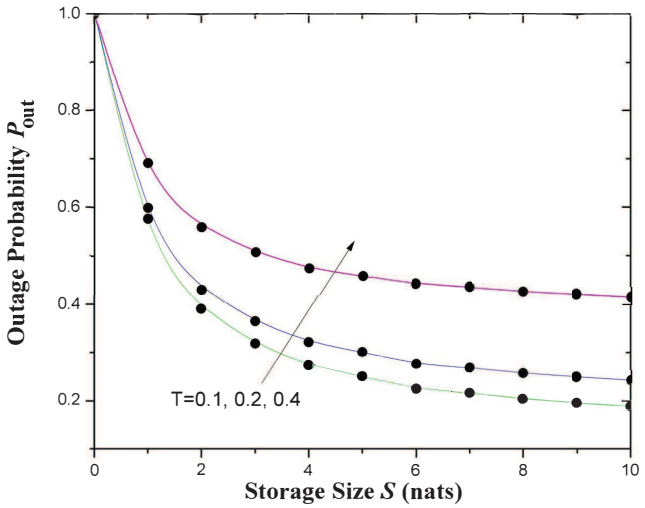

Figure 4: Outage probability versus the storage size S for varying target file bit-rate T. Solid lines and bullets correspond to the theoretical and simulated results, respectively.

B. Impact of Storage Size

Fig. 4 shows the relationship of the outage probability $\mathbb{P}_{\rm out}$ with the storage size S of the BSs. Notably, the storage capability of networks with caching is one of the most crucial parameters during the design. Obviously, the outage probability increases with the storage size, but decreases with the target bit-rate. In other words, the larger the target bit rate is set, the larger the outage probability will be.

In the same direction, in Fig. 5, the average delivery rate becomes higher with increasing storage size, but after a value of S, further increment is not beneficial, since all users content will be already available to the corresponding BSs. Especially, less target rate allows better coverage.

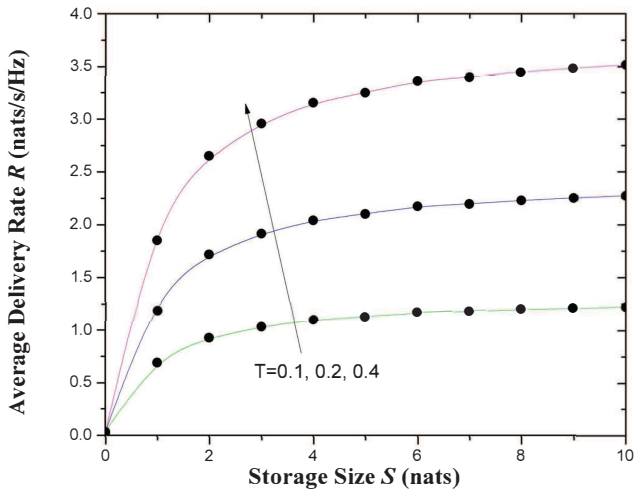

Figure 5: Average delivery rate versus the storage size S for varying target file bit-rate T. Solid lines and bullets correspond to the theoretical and simulated results, respectively.

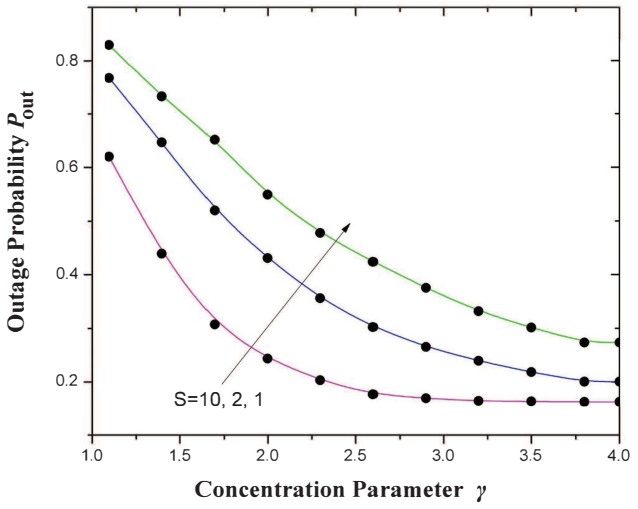

Figure 6: Outage probability versus the concentration parameter γ for varying storage size S. Solid lines and bullets correspond to the theoretical and simulated results, respectively.

C. Impact of Concentration Parameter

The variation of the concentration parameter, described by γ, is depicted in Fig. 6. Small γ means that a high quota of files is already at the intermediate nodes, i.e., many files are popular. Hence, the users do need to upload their files and significant outage is observed. Moreover, higher storage size allows more files to be uploaded in the BSs. As a result, it is likely that the contents of the users are already at the BSs and the users are inactive since they do not need to upload their contents.

The dependence of the average delivery rate R with the concentration parameter is provided by Fig. 7. Specifically, a

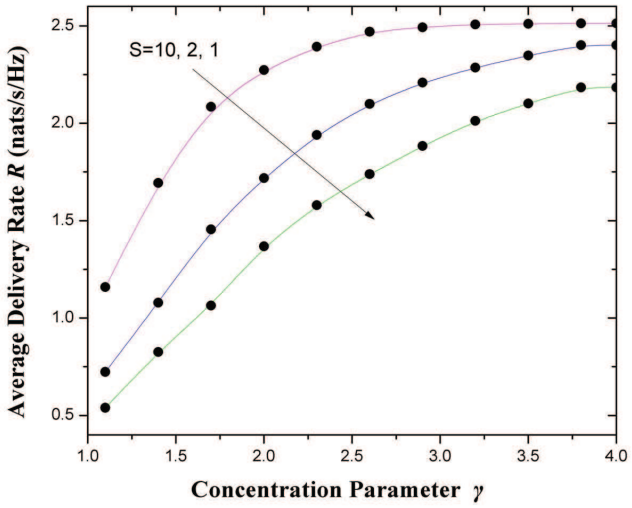

Figure 7: Average delivery rate versus the concentration parameter γ for varying storage size S. Solid lines and bullets correspond to the theoretical and simulated results, respectively.

high concentration parameter means that many files will be uploaded, and thus, the average rate increases.

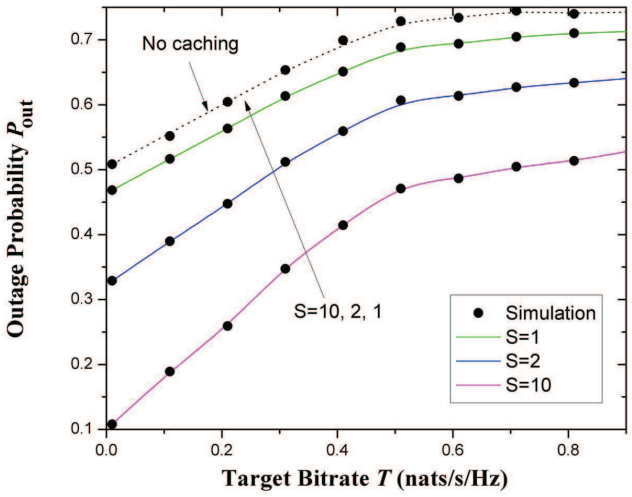

Figure 8: Outage probability versus the target bitrate T for varying storage size S. Solid lines and bullets correspond to the theoretical and simulated results, respectively, while the dotted line refers to the "No caching" scenario.

D. Impact of Target Bit-Rate

The target bit-rate T is another critical parameter that should be taken into account during the formation and study of the current architecture. In particular, Fig. 8 demonstrates the lines of the outage probability $\mathbb{P}_{\text{out}}$ versus the target T for $S = 1, 2$, and 10. Increasing the target rate, the outage probability increases, since less users are served. In addition, the performance is improved with increasing storate size because more content can be saved to the intermidiate nodes without the need to upload it at the backhaul.

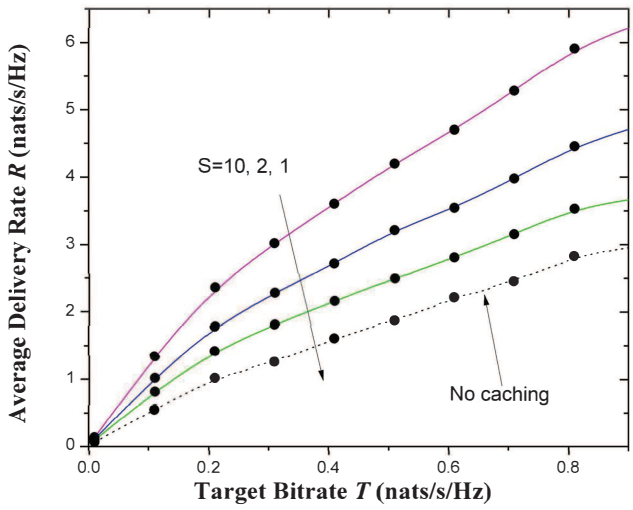

Figure 9: Average delivery rate versus the target bitrate T for varying storage size S. Solid lines and bullets correspond to the theoretical and simulated results, respectively, while the dotted line refers to the "No caching" scenario.

In a parallel avenue, Fig. 9 shows the increase in the performance with bigger storage capacities at the BSs, while it is apparent that a higher target rate results in a higher average delivery rate.

VIII. Conclusion

In this paper, we introduced the concept of caching in the uplink of a system with stochastically distributed massive MIMO BSs, where users upload their contents to servers through the BS by means of finite-rate backhaul links. In addition to significantly generalizing the state of the art cache-enabled PPP models to the uplink scenario, we enriched the uplink of the HetNet with the massive MIMO concept. Remarkably, it is the first work, where the caching nodes have a large number of antennas. Moreover, our approach considered imperfect CSI due to pilot contamination and channel aging. After deriving the DE of the SINR, we provided the outage probability and the average delivery rate. Our main purpose was to focus on fundamental parameters, being relevant to the caching design. Such parameters are the storage size of the serving BS and their target file rate. In particular, we demonstrated that by increasing the storage size, the performance of the system is improved, since the outage probability decreases and the average delivery rate increases. Furthermore, by increasing the target file bitrate, the majority of the users is not served. Hence, the outage probability increases. Overall, it was shown that the introduction of the notion of caching in the uplink enhances the system performance.

Appendix A
Useful Lemmas

Lemma 1 (Alzer's inequality [44])**.** *Assuming that h is a normalized gamma random variable with parameter N and*

a constant $\gamma >$, then the probability $\mathbb{P}\left(h < \gamma\right)$ can be tightly upper bounded by

$$\mathbb{P}\left(h < \gamma\right) < \left(1 - \mathrm{e}^{-\alpha\gamma}\right)^{N}, \tag{23}$$

where $\alpha = N\left(N!\right)^{-\frac{1}{N}}$.

Lemma 2 ([43, Lem. B.26]). *Let $\mathbf{A} \in \mathbb{C}^{N\times N}$ with uniformly bounded spectral norm (with respect to N). Consider $\mathbf{x}$ and $\mathbf{y}$, where $\mathbf{x}, \mathbf{y} \in \mathbb{C}^{N}$, $\mathbf{x} \sim \mathcal{CN}(\mathbf{0}, \mathbf{\Phi}_x)$ and $\mathbf{y} \sim \mathcal{CN}(\mathbf{0}, \mathbf{\Phi}_y)$, are mutually independent and independent of $\mathbf{A}$. Then, we have*

$$\frac{1}{N}\mathbf{x}^{\mathsf{H}}\mathbf{A}\mathbf{x} - \frac{1}{N}\operatorname{tr}\mathbf{A}\mathbf{\Phi}_x \xrightarrow[N\to\infty]{a.s.} 0 \tag{24}$$

$$\frac{1}{N}\mathbf{x}^{\mathsf{H}}\mathbf{A}\mathbf{y} \xrightarrow[N\to\infty]{a.s.} 0 \tag{25}$$

$$\mathbb{E}\left[\left|\left(\frac{1}{N}\mathbf{x}^{\mathsf{H}}\mathbf{A}\mathbf{x}\right)^{2} - \left(\frac{1}{N}\operatorname{tr}\mathbf{A}\mathbf{\Phi}_x\right)^{2}\right|\right] \xrightarrow[N\to\infty]{a.s.} 0 \tag{26}$$

$$\frac{1}{N^2}|\mathbf{x}^{\mathsf{H}}\mathbf{A}\mathbf{y}|^2 - \frac{1}{N^2}\operatorname{tr}\mathbf{A}\mathbf{\Phi}_x\mathbf{A}^{\mathsf{H}}\mathbf{\Phi}_y \xrightarrow[N\to\infty]{a.s.} 0. \tag{27}$$

Appendix B
Proof of Proposition 1

First, we divide both the numerator and the denominator of (16) by $\frac{1}{M^2}$. Then, we start with the numerator of the SINR. We insert in (16) the expression of the MRC decoder given by (15). We have[8]

$$\begin{aligned}\frac{1}{M^2}P_{lk}|\mathbf{q}_{llk}^{\mathsf{H}}\hat{\mathbf{h}}_{llk}|^2 &\overset{(a)}{=} \frac{\left(P_{lk}\beta_{llk}\right)^2}{\delta^2\left(\sum_j P_{jk}\beta_{ljk} + \frac{\sigma^2}{K}\right)^2}\frac{1}{M^2}|\mathbf{q}_{llk}|^4 \\ &\overset{(b)}{\asymp} \frac{1}{M}P_{\mathrm{t}}^2\delta^4\beta_{llk}^{2(1-\epsilon)},\end{aligned} \tag{28}$$

where (a) follows after substituting the estimated channel given in (15), while (b) is obtained by means of Lemma 2, since the covariance of $\mathbf{q}_{llk}$ is $\delta^2\left(\sum_j P_{ljk}\beta_{ljk} + \frac{\sigma^2}{K}\right)\mathbf{I}_M$. We continue with the first term in the denominator of the SINR including the estimation error. We have

$$\begin{aligned}&\frac{1}{M^2}P_{lk}|\mathbf{q}_{llk}^{\mathsf{H}}\tilde{\mathbf{e}}_{llk}|^2 \asymp \frac{1}{M^2}P_{lk}\delta^2 M\beta_{llk}\left(\sum_j P_{jk}\beta_{ljk} + \frac{\sigma^2}{K}\right) \\ &\times\left(1 - \delta^2\frac{P_{lk}\beta_{llk}}{\sum_j P_{jk}\beta_{ljk} + \frac{\sigma^2}{K}}\right) \\ &= P_{\mathrm{t}}^2\delta^2\beta_{llk}^{(1-\epsilon)}\frac{1}{M}\left(\sum_{j\neq l}\beta_{jjk}^{-\epsilon}\beta_{ljk} + \left(1-\delta^2\right)\beta_{llk}^{(1-\epsilon)} + \frac{\sigma^2}{P_{\mathrm{t}}K}\right)\end{aligned}$$

because the covariance of the estimation error is $\beta_{llk}\left(1 - \delta^2\frac{P_{lk}\beta_{llk}}{\sum_j P_{jk}\beta_{ljk}+\frac{\sigma^2}{K}}\right)\mathbf{I}_M$. The next step is to derive the second term in the denominator, which is written as

$$\begin{aligned}&\frac{1}{M^2}\sum_{(j,k')\neq(l,k)}|\sqrt{P_{jk'}}\mathbf{q}_{llk}^{\mathsf{H}}\mathbf{h}_{ljk'}|^2 \\ &= \frac{1}{M^2}\sum_{(j,k')\neq(l,k)}\delta^2|\sqrt{P_{jk'}}\mathbf{n}_{llk}^{\mathrm{tr}}\mathbf{h}_{ljk'} + \sum_{j'\neq l}\sqrt{P_{j'k}}\mathbf{h}_{lj'k}\mathbf{h}_{ljk'}|^2 \\ &\asymp \sum_{(j,k')\neq(l,k)}\delta^2\Big(\frac{\sigma^2}{MK}P_{\mathrm{t}}\beta_{jjk'}^{-\epsilon}\beta_{ljk'} + \frac{1}{M^2}\sum_{j'\neq l}P_{j'k}P_{jk'}|\mathbf{h}_{lj'k}\mathbf{h}_{ljk'}|^2\Big).\end{aligned} \tag{29}$$

When $k^{'} = k$ and $j^{'} = j$, the second term of the previous expression simplifies to

$$\begin{aligned}\frac{1}{M^2}P_{jk'}P_{j'k}|\mathbf{h}_{lj'k}\mathbf{h}_{ljk'}|^2 &= P_{jk}^2\frac{1}{M^2}\|\mathbf{h}_{ljk}\|^4 \\ &\asymp \frac{1}{M}P_{\mathrm{t}}^2\beta_{jjk}^{-2\epsilon}\beta_{ljk}^2,\end{aligned} \tag{30}$$

where we have used Lemma 2. In the case that $j^{'} \neq j$ and $k^{'} = k$ or $k^{'} \neq k$, we have

$$\begin{aligned}\frac{1}{M^2}P_{jk'}P_{j'k}\mathbf{h}_{lj'k}\mathbf{h}_{ljk'} &\asymp \frac{1}{M}P_{j'k}P_{jk'}\beta_{lj'k}\beta_{ljk'} \\ &= \frac{1}{M}P_{\mathrm{t}}^2\left(\beta_{j'j'k}\beta_{jjk'}\right)^{-\epsilon}\beta_{lj'k}\beta_{ljk'}.\end{aligned} \tag{31}$$

Thus, (29) becomes

$$\begin{aligned}&\sum_{(j,k')\neq(l,k)}\delta^2\Big(\frac{\sigma^2}{MK}MP_{\mathrm{t}}\beta_{jjk'}^{-\epsilon}\beta_{ljk'}\sum_{j'\neq l}P_{j'k}P_{jk'}\frac{1}{M^2}|\mathbf{h}_{lj'k}\mathbf{h}_{ljk'}|^2\Big) \\ &\asymp \sum_{(j,k')\neq(l,k)}\delta^2\frac{1}{M}\Big(\frac{\sigma^2}{K}P_{\mathrm{t}}\beta_{jjk'}^{-\epsilon}\beta_{ljk'} \\ &+ P_{\mathrm{t}}^2\beta_{jjk}^{-2\epsilon}\beta_{ljk}^2 + P_{\mathrm{t}}^2\left(\beta_{j'j'k}\beta_{jjk'}\right)^{-\epsilon}\beta_{lj'k}\beta_{ljk'}\Big).\end{aligned} \tag{32}$$

Regarding the term that includes the thermal noise, after applying Lemma 2 we have

$$\frac{1}{M^2}\|\mathbf{q}_{llk}\|^2\sigma^2 \asymp \frac{1}{M}P_{\mathrm{t}}\delta^2\left(\sum_j\beta_{jjk}^{-\epsilon}\beta_{ljk} + \frac{\sigma^2}{P_{\mathrm{t}}K}\right). \tag{33}$$

Appendix C
Proof of Theorem 1

The proof starts by finding first the conditional coverage probability on x as

$$\begin{aligned}\tilde{\mathbb{P}}\Big(\overline{\mathrm{SINR}}_{llk} > \tilde{T}, f\notin\Delta_{b_0}\Big) &= \mathbb{E}_x\Big[\tilde{\mathbb{P}}\left(\overline{\mathrm{SINR}}_{llk} > \tilde{T}\right)|x\Big] \\ &\times \mathbb{E}_x\left[\mathbb{P}\left(f\notin\Delta_{b_0}\right)|x\right].\end{aligned} \tag{34}$$

Hence, we focus on the derivation of $\tilde{\mathbb{P}}\left(\overline{\mathrm{SINR}}_{llk} > \tilde{T}|x\right)$. Specifically, we propose an approximation for the out-of-cell interference, described by W_{jk} for all users in each cell, i.e., $k \in [1, K]$. This approximation will allow the decoupling of the correlated terms and will result in a tractable evaluation of

[8]Let a_n and b_n two infinite sequences. $a_n \asymp b_n$ denotes the equivalence relation $a_n - b_n \xrightarrow[N\to\infty]{\text{a.s.}} 0$.

$\mathbb{P}_c$. Specifically, by approximating the out-of-cell interference with its mean, we have [45]

$$\begin{aligned}\sum_{j\neq l} W_{jk} &= \sum_{j\neq l} \beta_{jjk}^{-\epsilon}\beta_{ljk} \\ &\approx \mathbb{E}\left[\sum_{j\neq l} \beta_{jjk}^{-\epsilon}\beta_{ljk}\right] \\ &= \mathbb{E}\left[\sum_{j\neq l} \mathbb{E}\left[\beta_{jjk}^{-\epsilon}\right]\beta_{ljk}\right] \qquad (35)\\ &= \lambda_{\mathrm{b}} C^{-\epsilon}\left(\pi\lambda_{\mathrm{b}}\right)^{-\frac{\alpha\epsilon}{2}}\Gamma^{\alpha}\left(\frac{\epsilon}{2}+1\right)\mathbb{E}\left[\sum_{j\neq l}\beta_{ljk}\right] \\ &= \frac{2C^{1-\epsilon}\left(\pi\lambda_{\mathrm{b}}\right)^{\frac{\alpha(1-\epsilon)}{2}}}{\alpha-2}\Gamma^{\alpha}\left(\frac{\epsilon}{2}+1\right), \qquad (36)\end{aligned}$$

where in (35), we made the following substitution

$$\begin{aligned}\mathbb{E}\left[\beta_{jjk}^{-\epsilon}\right] &= C^{-\epsilon}\left(\mathbb{E}\left[r_{ljk}^{\epsilon}\right]\right)^{\alpha} \\ &= \left(\pi\lambda_{\mathrm{b}}\right)^{-\frac{\alpha\epsilon}{2}}\Gamma^{\alpha}\left(\frac{\epsilon}{2}+1\right).\end{aligned}$$

Moreover, (36) is obtained by means of the Campbell's theorem [30] and the exclusion ball model, described in Sec. II as

$$\begin{aligned}\mathbb{E}\left[\sum_{j\neq l}\beta_{ljk}\right] &= 2\pi\lambda_{\mathrm{b}} C\int_{R_e}^{\infty} x^{-\alpha}x\mathrm{d}x \\ &= \frac{2C\left(\lambda_{\mathrm{b}}\right)^{\frac{\alpha}{2}}\pi\lambda_{\mathrm{b}}}{\alpha-2}. \qquad (37)\end{aligned}$$

In a similar way, we can write

$$\sum_{j\neq l} W_{jk}^2 = \frac{2C^{2(1-\epsilon)}\left(\pi\lambda_{\mathrm{b}}\right)^{\alpha(1-\epsilon)}}{\alpha-1}\Gamma^{\alpha}\left(\epsilon+1\right). \qquad (38)$$

The approximations (36) and (38) allow the simplification of the SINR, conditioned on $r_{jjk}=x$, by its approximate

$$\begin{aligned}\overline{\mathrm{SINR}} \approx \Big(&D_1\left(\pi\lambda_B x^2\right)^{\alpha(1-\epsilon)} + D_2\left(\pi\lambda_B x^2\right)^{2\alpha(1-\epsilon)}\sum_{k'\neq k} r_{jjk}^{\alpha(1-\epsilon)} \\ &+ D_3\left(\pi\lambda_B x^2\right)^{2\alpha(1-\epsilon)} + D_4\left(\pi\lambda_B x^2\right)^{2\alpha(1-\epsilon)} + D_6\Big)^{-1}, \qquad (39)\end{aligned}$$

where in (39) the variables $D_i \in 1,\ldots,4,6$ are defined in Theorem 1. Conditioned on $r_{jjk}=x$, we obtain the approximate distribution of the SINR, given by (40), after substituting its expression from (39). In (41), we have considered the dummy variable $\tilde{g}$, having unit mean and shape parameter N, in order to approximate the constant number one. Actually, this approximation becomes tighter as N goes to infinity [44], since $\lim_{y\to\infty}\frac{y^y x^{y-1}\mathrm{e}^{-yx}}{\Gamma(y)} = \delta\left(x-1\right)$ with $\delta\left(x\right)$ being Dirac's delta function. In (42), we have applied Alzer's inequality (see Lemma 1), where $\eta = N\left(N!\right)^{-\frac{1}{N}}$, while afterwards, in (43), we have used the Binomial theorem. Next, (44) is obtained by assuming that y is a Rayleigh random variable. In (45), we set $u=\lambda_B\pi y^2$, and we take into account the approximation $\exp\left(-x\right)\approx\frac{1}{1+x}$, in order to make the numerical integration faster. Finally, given that x is a Rayleigh random variable, we obtain the uplink SINR distribution as

$$\begin{aligned}\mathbb{E}_x\left[\tilde{\mathbb{P}}\left(\overline{\mathrm{SINR}}_{llk}>\tilde{T}\right)|x\right] = &\int_0^{\infty}\tilde{\mathbb{P}}\left(\overline{\mathrm{SINR}}>\tilde{T}|r_{jjk}=x\right) \\ &\times e^{-\pi\lambda_{\mathrm{B}}x^2}2\pi\lambda_{\mathrm{B}}x\mathrm{d}x. \qquad (46)\end{aligned}$$

The derivation of the second term of (34) is straightforward. In particular, we assume that all the BSs cache the same amount of files (storage size), while the cache hit probability is independent of the distance $r_{jjk}=x$. Thus, we have

$$\mathbb{E}_x\left[\mathbb{P}(f\notin\Delta_{b_0}|x)\right] = 1-\int_0^{\frac{S}{L}} f_{\mathrm{pop}}(f,\mu,\gamma)\mathrm{d}f. \qquad (47)$$

Inserting (46) and (47) into (34) and after some algebraic manipulations, we obtain the coverage probability. The proof is concluded by substituting the coverage probability $\tilde{\mathbb{P}}\left(\overline{\mathrm{SINR}}_{llk}>\tilde{T}, f\notin\Delta_{b_0}\right)$ in (18).

Appendix D
Proof of Theorem 2

The average delivery rate $\bar{R}=\mathbb{E}\left(R\right)$ is obtained by applying the expectation operator over both the fading distribution and the PPP. In particular, we have

$$\begin{aligned}\bar{R} &= \mathbb{E}\left[R\right] \\ &= \psi\mathbb{E}\left[\tilde{\mathbb{P}}\left(\ln\left(1+\overline{\mathrm{SINR}}_{llk}\right)>T\right)\right. \\ &\times\left.\left(\,T\mathbb{P}(f\notin\Delta_{b_0})+C\left(\lambda_{\mathrm{B}}\right)\mathbb{P}(f\in\Delta_{b_0}\,)\right)\right] \qquad (48)\\ &= \psi\mathbb{E}\left[\tilde{\mathbb{P}}\left(\ln\left(1+\overline{\mathrm{SINR}}_{llk}\right)>T|x\right)\right] \\ &\times\left(\,\mathbb{E}\left[T\mathbb{P}(f\notin\Delta_{b_0}|x)\right]+\mathbb{E}\left[C\left(\lambda_{\mathrm{B}}\right)\mathbb{P}(f\in\Delta_{b_0}|x)\right]\right. \qquad (49)\\ &= \psi\mathcal{I}_1\left(\mathcal{I}_2+\mathcal{I}_3\right), \qquad (50)\end{aligned}$$

where (48) follows by applying the definition described by (20), while (49) is obtained because of the independence between the different events and the property of linearity of the expectation operator. The derivation of $\bar{R}$ continues with the substitution of $\mathcal{I}_1$, which is basically given by (21) after substituting $\tilde{T}$ with T. Moreover, given that the cache hit probability does not depend on x, we have

$$\mathcal{I}_2 = T\int_0^{\frac{S}{L}} f_{\mathrm{pop}}(f,\mu,\gamma)\mathrm{d}f, \qquad (51)$$

while $\mathcal{I}_3$ is obtained by multiplying (47) with $C\left(\lambda_{\mathrm{B}}\right)$. After appropriate substitutions in (50), the proof is concluded. □

Acknowledgement

The authors would like to express their gratitude to Dr. E. Baştuğ for his help and support in making this work possible.

References

[1] I. Cisco Visual Networking, "Global mobile data traffic forecast update 2014–2019. White paper c11-520862," *Available on http://www.cisco. com/c/en/us/solutions/collateral/serviceprovider/visual-networking-index-vni/white_paper_c11-520862. html.*

[2] J. G. Andrews *et al.*, "What will 5G be?" *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1065–1082, June 2014.

$$\tilde{\mathbb{P}}\Big(\overline{\mathrm{SINR}} > \tilde{T}|r_{jjk} = x\Big)$$

$$\approx \tilde{\mathbb{P}}\Big(1 > \tilde{T}\delta^{-2}\Big(D_1\big(\pi\lambda_B x^2\big)^{\alpha(1-\epsilon)} + D_2\big(\pi\lambda_B x^2\big)^{2\alpha(1-\epsilon)}\sum_{k'\neq k} r_{jjk'}^{\alpha(1-\epsilon)} + D_3\big(\pi\lambda_B x^2\big)^{2\alpha(1-\epsilon)} + D_4\big(\pi\lambda_B x^2\big)^{2\alpha(1-\epsilon)} + D_6\Big)\Big) \tag{40}$$

$$\approx \tilde{\mathbb{P}}\Big(\tilde{g} > \tilde{T}\delta^{-2}\Big(D_1\big(\pi\lambda_B x^2\big)^{\alpha(1-\epsilon)} + D_2\left(\pi\lambda_B x^2\right)^{2\alpha(1-\epsilon)}\sum_{k'\neq k} r_{jjk'}^{\alpha(1-\epsilon)} + D_3\big(\pi\lambda_B x^2\big)^{2\alpha(1-\epsilon)} + D_4\big(\pi\lambda_B x^2\big)^{2\alpha(1-\epsilon)} + D_6\Big)\Big) \tag{41}$$

$$\approx 1 - \mathbb{E}\Bigg[\Bigg(1 - \exp\Bigg(-\eta\tilde{T}\delta^{-2}\Big(D_1\left(\pi\lambda_B x^2\right)^{\alpha(1-\epsilon)} + D_2\left(\pi\lambda_B x^2\right)^{2\alpha(1-\epsilon)}\sum_{k'\neq k} r_{jjk'}^{\alpha(1-\epsilon)} + D_3\left(\pi\lambda_B x^2\right)^{2\alpha(1-\epsilon)} + D_4\left(\pi\lambda_B x^2\right)^{2\alpha(1-\epsilon)} + D_6\Big)\Bigg)\Bigg)^N\Bigg] \tag{42}$$

$$= \sum_{n=1}^{N}\binom{N}{n}(-1)^{n+1}\,\mathbb{E}\Bigg[\exp\Big(-n\eta\tilde{T}\delta^{-2}\Big(D_1\left(\pi\lambda_B x^2\right)^{\alpha(1-\epsilon)} + D_2\left(\pi\lambda_B x^2\right)^{2\alpha(1-\epsilon)}\sum_{k'\neq k} y_{k'}^{\alpha(1-\epsilon)} + D_3\left(\pi\lambda_B x^2\right)^{2\alpha(1-\epsilon)} + D_4\left(\pi\lambda_B x^2\right)^{2\alpha(1-\epsilon)} + D_6\Big)\Big)\Bigg] \tag{43}$$

$$= \sum_{n=1}^{N}\binom{N}{n}(-1)^{n+1}\exp\Bigg(-n\eta\tilde{T}\delta^{-2}\Big(D_1\left(\pi\lambda_B x^2\right)^{\alpha(1-\epsilon)} + D_3\left(\pi\lambda_B x^2\right)^{2\alpha(1-\epsilon)} + D_4\left(\pi\lambda_B x^2\right)^{2\alpha(1-\epsilon)} + D_6\Big)\Bigg) \times \int_0^{\infty}\exp\Big(D_2\left(\pi\lambda_B x^2\right)^{2\alpha(1-\epsilon)} y^{\alpha(1-\epsilon)-\lambda_B\pi y^2}\Big)^{K-1}\lambda_B\pi y\mathrm{d}y \tag{44}$$

$$= \sum_{n=1}^{N}\binom{N}{n}(-1)^{n+1}\exp\Bigg(-n\eta\tilde{T}\delta^{-2}\Big(D_1\left(\pi\lambda_B x^2\right)^{\alpha(1-\epsilon)} + D_3\left(\pi\lambda_B x^2\right)^{2\alpha(1-\epsilon)} + D_4\left(\pi\lambda_B x^2\right)^{2\alpha(1-\epsilon)} + D_6\Big)\Bigg) \times \Bigg(1 - D_2\left(\pi\lambda_B x^2\right)^{2\alpha(1-\epsilon)}\int_0^{\infty}\frac{\mathrm{e}^{-u}\mathrm{d}u}{1 + D_2\left(\pi\lambda_B x^2\right)^{2\alpha(1-\epsilon)}u^{-\frac{\alpha}{2}(1-\epsilon)}}\Bigg)^{K-1} \tag{45}$$

[3] J. G. Andrews, H. Claussen, M. Dohler, S. Rangan, and M. C. Reed, "Femtocells: Past, present, and future," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 3, pp. 497–508, Apr. 2012.

[4] J. G. Andrews, F. Baccelli, and R. K. Ganti, "A tractable approach to coverage and rate in cellular networks," *IEEE Trans. Commun.*, vol. 59, no. 11, pp. 3122–3134, Nov. 2011.

[5] P. Madhusudhanan, J. G. Restrepo, Y. Liu, T. X. Brown, and K. R. Baker, "Multi-tier network performance analysis using a shotgun cellular system," in *IEEE Global Telecommunications Conference (GLOBECOM 2011)*, 2011, pp. 1–6.

[6] H.-S. Jo, Y. J. Sang, P. Xia, and J. G. Andrews, "Heterogeneous cellular networks with flexible cell association: A comprehensive downlink SINR analysis," *IEEE Trans. Wireless Commun.*, vol. 11, no. 10, pp. 3484–3495, Oct. 2012.

[7] S. Mukherjee, "Distribution of downlink SINR in heterogeneous cellular networks," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 3, pp. 575–585, Apr. 2012.

[8] C. B. Sankaran, "Data offloading techniques in 3GPP Rel-10 networks: A tutorial," *IEEE Commun. Mag.*, vol. 50, no. 6, pp. 46–53, June 2012.

[9] R. W. Heath, M. Kountouris, and T. Bai, "Modeling heterogeneous network interference using Poisson point processes," *IEEE Trans. Signal Process.*, vol. 61, no. 16, pp. 4114–4126, Aug. 2013.

[10] H. S. Dhillon, M. Kountouris, and J. G. Andrews, "Downlink MIMO HetNets: Modeling, ordering results and performance analysis," *IEEE Trans. Wireless Commun.*, vol. 12, no. 10, pp. 5208–5222, Oct. 2013.

[11] A. K. Papazafeiropoulos and T. Ratnarajah, "Downlink MIMO HCNs with residual transceiver hardware impairments," *IEEE Commun. Lett.*, vol. 20, no. 10, pp. 2023–2026, Oct. 2016.

[12] M. Kountouris and J. G. Andrews, "Downlink SDMA with limited feedback in interference-limited wireless networks," *IEEE Trans. Wireless Commun.*, vol. 11, no. 8, pp. 2730–2741, Aug. 2012.

[13] A. Papazafeiropoulos and T. Ratnarajah, "Towards a realistic assessment of multiple antenna HCNs: Residual additive transceiver hardware impairments and channel aging," *IEEE Trans. Veh. Technol.*, vol. 66, no. 10, pp. 9061–9073, Oct. 2017.

[14] T. D. Novlan, H. S. Dhillon, and J. G. Andrews, "Analytical modeling of uplink cellular networks," *IEEE Trans. Wireless Commun.*, vol. 12, no. 6, pp. 2669–2679, June 2013.

[15] T. Bai and R. W. Heath, "Analyzing uplink SINR and rate in massive MIMO systems using stochastic geometry," *IEEE Trans. Commun.*, vol. 64, no. 11, pp. 4592–4606, Nov. 2016.

[16] H. Q. Ngo, E. Larsson, and T. Marzetta, "Energy and spectral efficiency of very large multiuser MIMO systems," *IEEE Trans. Commun.*, vol. 61, no. 4, pp. 1436–1449, Apr. 2013.

[17] L. Lu, G. Y. Li, A. L. Swindlehurst, A. Ashikhmin, and R. Zhang, "An overview of massive MIMO: Benefits and challenges," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 742–758, Oct. 2014.

[18] T. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, Nov. 2010.

[19] E. Björnson, J. Hoydis, M. Kountouris, and M. Debbah, "Massive MIMO systems with non-ideal hardware: Energy efficiency, estimation, and capacity limits," *IEEE Trans. Inf. Theory*, vol. 60, no. 11, pp. 7112–7139, Nov. 2014.

[20] E. Björnson, M. Matthaiou, and M. Debbah, "Massive MIMO with non-ideal arbitrary arrays: Hardware scaling laws and circuit-aware design," *IEEE Trans. Wireless Commun.*, vol. 14, no. 8, pp. 4353–4368, Aug. 2015.

[21] A. K. Papazafeiropoulos and T. Ratnarajah, "Deterministic equivalent performance analysis of time-varying massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 14, no. 10, pp. 5795–5809, Oct. 2015.

[22] A. K. Papazafeiropoulos, "Impact of user mobility on optimal linear receivers in cellular networks," in *IEEE International Conference on Communications (ICC)*, 2015, pp. 2239–2244.

[23] ——, "Impact of general channel aging conditions on the downlink performance of massive MIMO," *IEEE Trans. Veh. Technol.*, vol. 66, no. 2, pp. 1428–1442, Feb. 2017.

[24] E. Baştuğ, M. Bennis, and M. Debbah, "Living on the edge: The role of proactive caching in 5G wireless networks," *IEEE Commun. Mag.*, vol. 52, no. 8, pp. 82–89, May 2014.

[25] H. Liu, Z. Chen, X. Tian, X. Wang, and M. Tao, "On content-centric wireless delivery networks," *IEEE Wireless Commun.*, vol. 21, no. 6, pp. 118–125, Dec. 2014.

[26] Z. Zhao, M. Peng, Z. Ding, W. Wang, and H. V. Poor, "Cluster content caching: An energy-efficient approach to improve quality of service in cloud radio access networks," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 5, pp. 1207–1221, May 2016.

[27] E. Baştuğ, M. Bennis, M. Kountouris, and M. Debbah, "Cache-enabled

small cell networks: Modeling and tradeoffs," *EURASIP J. Wireless Commun. and Net.*, vol. 2015, no. 1, p. 41, Dec. 2015.
[28] E. Björnson, L. Sanguinetti, and M. Kountouris, "Deploying dense networks for maximal energy efficiency: Small cells meet massive MIMO," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 4, pp. 832–847, Apr. 2016.
[29] A. Tang, J. Sun, and K. Gong, "Mobile propagation loss with a low base station antenna for NLOS street microcells in urban area," in *IEEE VTS 53rd Vehicular Technology Conference, Spring 2001.*, vol. 1. IEEE, 2001, pp. 333–336.
[30] S. N. Chiu, D. Stoyan, W. S. Kendall, and J. Mecke, *Stochastic Geometry and its Applications*. John Wiley & Sons, 2013.
[31] N. Liang, W. Zhang, and C. Shen, "An uplink interference analysis for massive MIMO systems with MRC and ZF receivers," in *IEEE Wireless Commun. and Net. Conf. (WCNC), 2015*, pp. 310–315.
[32] F. Baccelli and B. Błaszczyszyn, *Stochastic Geometry and Wireless Networks: Volume I Theory*. NoW Publishers, 2009.
[33] H. ElSawy, E. Hossain, and M. Haenggi, "Stochastic geometry for modeling, analysis, and design of multi-tier and cognitive cellular wireless networks: A survey," *IEEE Commun. Surveys Tuts.*, vol. 15, no. 3, pp. 996–1019, July 2013.
[34] X. Zhang and J. G. Andrews, "Downlink cellular network analysis with multi-slope path loss models," *IEEE Trans. Commun.*, vol. 63, no. 5, pp. 1881–1894, May 2015.
[35] Cisco. Cisco visual networking index: Global mobile data traffic forecast update, 2016–2021 white paper. [Online]. Available: https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/mobile-white-paper-c11-520862.html
[36] K. Mardia and P. Zemroch, "Algorithm AS 86: The von Mises distribution function," *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, vol. 24, no. 2, pp. 268–272, 1975.
[37] K. E. Baddour and N. C. Beaulieu, "Autoregressive modeling for fading channel simulation," *IEEE Trans. Wireless Commun.*, vol. 4, no. 4, pp. 1650–1662, July 2005.
[38] M. Vu and A. Paulraj, "On the capacity of MIMO wireless channels with dynamic CSIT," *IEEE J. Select. Areas Commun.*, vol. 25, no. 7, pp. 1269–1283, Sep. 2007.
[39] W. C. Jakes and D. C. Cox, *Microwave mobile communications*. Wiley-IEEE Press, 1994.
[40] R. Couillet and M. Debbah, *Random Matrix Methods for Wireless Communications*. Cambridge University Press, 2011.
[41] A. W. van der Vaart, *Asymptotic statistics (Cambridge series in statistical and probabilistic mathematics)*. New York: Cambridge University Press, 2000.
[42] S. Wagner, R. Couillet, M. Debbah, and D. Slock, "Large system analysis of linear precoding in correlated MISO broadcast channels under limited feedback," *IEEE Trans. Inf. Theory*, vol. 58, no. 7, pp. 4509–4537, July 2012.
[43] Z. Bai and J. W. Silverstein, *Spectral Analysis of Large Dimensional Random Matrices*. Springer, 2010, vol. 20.
[44] H. Alzer, "On some inequalities for the incomplete gamma function," *Mathematics of Computation of the American Mathematical Society*, vol. 66, no. 218, pp. 771–778, 1997.
[45] R. K. Mungara, D. Morales-Jimenez, and A. Lozano, "System-level performance of interference alignment," *IEEE Trans. Wireless Commun.*, vol. 14, no. 2, pp. 1060–1070, Feb. 2015.