5G Massive MIMO Architectures: Self-Backhauled Small Cells versus Direct Access

Andrea Bonfante, Lorenzo Galati Giordano, David López-Pérez, Adrian García-Rodríguez, Giovanni Geraci, Paolo Baracca, M. Majid Butt, and Nicola Marchetti

Abstract—In this paper, we focus on one of the key technologies for the fifth-generation wireless communication networks, massive multiple-input-multiple-output (mMIMO), by investigating two of its most relevant architectures: 1) to provide in-band backhaul for the ultra-dense network (UDN) of self-backhauled small cells (SCs), and 2) to provide direct access (DA) to user equipments (UEs). Through comprehensive 3GPP-based system-level simulations and analytical formulations, we show the end-to-end UE rates achievable with these two architectures. Differently from the existing works, we provide results for two strategies of self-backhauled SC deployments, namely random and ad-hoc, where in the latter SCs are purposely positioned close to UEs to achieve line-of-sight (LoS) access links. We also evaluate the optimal backhaul and access time resource partition due to the in-band self-backhauling (s-BH) operations. Our results show that the ad-hoc deployment of self-backhauled SCs closer to the UEs with optimal resource partition and with directive antenna patterns, provides rate improvements for cell-edge UEs that amount to 30% and tenfold gain, as compared to mMIMO DA architecture with pilot reuse 3 and 1 reuse, respectively. On the other hand, mMIMO s-BH underperforms mMIMO DA above the median value of the UE rates when the effect of pilot contamination is less severe, and the LoS probability of the DA links improves.

Index Terms—5G mobile communication, massive MIMO, wireless backhaul, small cell deployment, network capacity.

I. INTRODUCTION

Fifth-generation (5G) wireless communication systems are expected to support a 1000× increase in capacity compared to existing networks [2]. Meeting this gargantuan target will require mobile network operators (MNOs) to leverage new technologies, such as massive multiple-input-multiple-output (mMIMO), and network densification using small cells base stations (SCs) [3], [4]. More specifically, mMIMO consists of deploying a large number of antennas at the base station (BS) to enable the simultaneous transmission to and from a multitude of terminals which are spatially multiplexed in the same time-frequency resources [3], while network densification considers the deployment of a large number of SCs up to the point of having one SC per user equipment (UE) [4]. A further step towards the practical ultra-dense SC deployment is represented by self-backhauling (s-BH), defined as the capability for a BS to enable in the same spectrum resources both the access and the backhaul transmissions. Self-backhauling could drive the tight integration between mMIMO and ultra-dense SC deployment luring MNOs with the potential of achieving the desired capacity boost at a contained investment [5]. Indeed, exploiting the large number of spatial degrees-of-freedom (DoF) available with mMIMO to provide sub-6 GHz in-band wireless backhauling to SCs offers multiple advantages to MNOs: i) avoiding deployment of an expensive wired backhaul infrastructure, ii) availing of more flexibility in the deployment of SCs, and iii) not having to purchase additional licensed spectrum, as in the case of out-of-band wireless backhauling [6].

Those advantages motivated the Third Generation Partnership Project (3GPP) to include in the 5G New Radio (NR) Release 15 a new study item, which focuses on Integrated Access and Backhaul (IAB) network architectures — also referred to as self-backhauling networks in the literature. In [7], 3GPP provides a list of use cases, in both the sub-6 GHz and above-6 GHz spectrum bands, and network architecture requirements for the NR backhauling functionalities coupled with the radio access network (RAN) technology.

A. Background and Related Work

Several works focused on millimeter wave (mmWave) s-BH networks [8], [9], which offer wide bandwidth channels to accommodate multiple backhaul and access links simultaneously. At the same time, various research efforts considered sub-6 GHz s-BH networks in a heterogeneous network (HetNet) environment [10]–[12], which is more suitable to provide wide-area coverage through conventional macro-cells, and use self-backhauled SCs to further boost the network capacity. However, due to the scarcity of spectrum in the bands below 6 GHz, the bandwidth splitting required to serve multiple backhaul links and the inter-tier interference caused by co-channel access and backhaul operations may turn out to be
a significant impediment to the potential adoption of s-BH in sub-6 GHz HetNets [12]. In [10], [11], the authors tackled the problem of resource allocation (such as transmission power and time-frequency resources) of s-BH networks. In [12], the authors considered a full-duplex (FD) technique, which on the one hand avoids the partitioning of the spectrum between backhaul and access, but on the other hand, it requires to study the problem of the self-interference due to the possible bi-directional transmission.

Moreover, works such as [13], [14], considered macro BSs equipped with mMIMO to enhance the backhaul link capacity and simultaneously serve SCs and UEs. In [13], the authors studied the UE data-rate performance as a function of the distance between the mMIMO-BS and the s-BH SCs. However, they considered a simplified single-cell scenario without inter-cell interference between SCs. In [14], [15], the authors used stochastic geometry to derive the rate coverage probability and compute the optimal proportion of in-band and out-of-band FD SCs in the network which maximizes the UE rates, and energy efficiency, respectively. Finally, the authors in [16] investigated an optimization approach to maximize the sum-rate of the UEs under capacity constrained backhaul and by considering the length of the pilot sequences used for the channel state information (CSI) acquisition.

B. Motivation and Contribution

In this paper, we analyze the end-to-end UE performance of mMIMO s-BH architecture by means of theoretical analysis and 3GPP-based system-level simulations when compared to mMIMO Direct Access (DA), where mMIMO-BSs are solely dedicated to serving UEs in the absence of SCs [17]. We consider a realistic multi-cell setup [18], where mMIMO-BSs provide sub-6 GHz backhauling to a plurality of half-duplex (HD) SCs overlaying the macro cellular area. In these HD systems, a s-BH network entails sharing time-and-frequency resources between radio access and backhaul links. We use 3GPP-based system level simulations up to the point of one SC deployed for each UE, whereas we use the analytical framework when the SC density is much larger than the UE density. In this regime, numerical simulations become impractical due to the large amount of time required and the high computational complexity involved. Two different strategies of self-backhauled SC deployments are considered as illustrated in Fig. 1. We analyze a random deployment – where SCs are uniformly distributed over a geographical area –, and an ad-hoc deployment – where SCs are purposely positioned close to UEs to achieve line-of-sight (LoS) access links. We would like to stress that the particular choice of the ad-hoc SC deployment serves two purposes: a) to study the performance gains that can be obtained by the MNOs through a tailored SCs deployment, and b) to provide an upper bound of the system access performance obtained when the SC locations are coupled to the UE locations.

The contributions of the paper are as follows:

1) We provide 3GPP-based system-level simulations results on the performance of the achievable UE data-rates in mMIMO based wireless in-band s-BH with random and ad-hoc SC deployments. Differently from the previous works [13]–[16], our work accounts for a pathloss model with LoS and Non-LoS (NLoS) transitions in both backhaul and access links. The closest similar work is [13], which also compares the performance of mMIMO s-BH and mMIMO DA architectures by considering channel models based only on pathloss, but not accounting for the pilot contamination effect on the signal-to-interference-plus-noise ratio (SINR) when mMIMO is used. On the contrary, we show that both the pathloss models incorporating LoS and NLoS transmissions and the pilot contamination effect severely impact the inter-cell interference modeling and the system performance, making our results more accurate and realistic than those presented in [13].

2) We provide an analytical model for evaluating the average data-rate of the backhaul and access links. We adapt the expressions proposed in [19], [20] to model the mMIMO backhaul links. Specifically, we account for the effects of the antenna directionality and sectorization, and the effect of the beamforming gain due to the mMIMO precoding. In the access link formulation, we account for the density of active SCs, which matches the numerical results obtained by simulations. We also employ the analytical framework to show that, due to the over provisioning of self-backhauled SCs, random deployment requires thousands of SCs to achieve the same performance of the ad-hoc deployment.

3) We explain in details the different factors playing a key role in the 3GPP-based system-level results. Overall, the insights from these results can guide the deployment of future 5G access network.

The remainder of this paper is organized as follows. Section II introduces the system model on which the analysis is based; Section III presents the downlink (DL) SINR and rate expressions of the backhaul and access links; Section IV provides the analytical signal-to-interference ratio (SIR) and average rate expressions of backhaul and access; Section V presents the numerical results, and VI summarizes the key findings. The following notation is used throughout the paper. Capital and lower-case bold letters denote matrices and vectors, respectively, while \([\cdot]^*, [\cdot]^T\) and \([\cdot]^H\) denote conjugate, transpose, and conjugate transpose, respectively. \(|\cdot|\) indicates the absolute value. The notation \(\mathbb{E}[\cdot]\) indicates the expectation with respect to a random variable (RV). We use \(\mathbb{C}\) and \(\mathbb{R}\) to denote complex- and real-valued numbers, respectively.

II. SYSTEM MODEL

We focus on the study of the DL performance for an hexagonal grid of mMIMO-BSs equipped with a large number of antennas \(M\) and providing wireless backhaul links to (a) randomly deployed self-backhauled SCs, (b) ad-hoc deployed self-backhauled SCs, or (c) directly serving UEs, as illustrated in Figs. 1(a), 1(b) and 1(c), respectively. Throughout the paper, we assume to use sub-6 GHz frequencies, which can provide backhaul connections with a single hop to all the SCs deployed in the coverage area of the macro cell [5].

For the mMIMO-BSs, we consider the hexagonal model since it represents a planned scenario of deployment where
the BS locations are optimized to improve the SINR performance. This deployment strategy is in line with the 3GPP methodology described in [18], which we take as a reference throughout the paper.

For the SCs, we differentiate the modeling between the random and the ad-hoc cases. In the first case, we model the SC locations with a homogeneous spatial poisson-point-process (SPPP) distribution, which we refer to as SPPP, since it represents an unplanned scenario of deployment, where the SCs are distributed according to a uniform distribution. This approach is in line with the works such as [21], [22], and also with the 3GPP methodology [18]. In the second case, we consider the SC locations coupled to the UE point process, and we position the SCs at a fixed distance $d$ with respect to the UE locations.

While the random deployment of SCs is agnostic to the UE positions, the ad-hoc deployment of SCs requires knowing the position of the UEs before making the deployment. This information can be obtained through predictions, for example, considering the behavior of UEs, who are likely to return to the same place at the same time of day.

### A. Macro Cell, Small Cell and User Topology

We denote by $i$, $l$, and $k$ the mMIMO-BS in the $i$-th sector, the SC, and the UE, respectively. $I$ represents the set of mMIMO-BSs deployed in the network. $L_i$ and $L_{i'}$ represent the set of SCs deployed per sector and connected to the $i$-th and $i'$-th mMIMO-BS, respectively, which provides the largest reference signal received power (RSRP). $L_i$ and $L_{i'}$ denote the number of SCs in the sets $L_i$ and $L_{i'}$, respectively. Furthermore, we denote by $K_l$ the number of UEs in the set $K_l$ randomly and uniformly distributed over the area covered by each sector. We assume that each single-antenna UE is connected with the SC (in the s-BH network), or with the mMIMO-BS (in the DA network) that provides the largest RSRP [18]. Therefore each SC serves $K_l$ UEs in the s-BH network.

Three different network deployments are presented in the following:

(a) **s-BH architecture with random deployment:** Self-backhauled SCs are randomly and uniformly distributed over the mMIMO-BS geographical area, as shown in Fig. 1(a). This scenario is used as a baseline, and follows the set of parameters specified by the 3GPP in [18] to evaluate the relay scenario. More precisely, we consider the UE and SC antenna heights fixed at 1.5 and 5 meters above the ground, respectively, and channel models for the 3GPP Case 1 Relay scenario.

(b) **s-BH architecture with ad-hoc deployment:** Self-backhauled SCs are positioned targeting nearby UE locations\(^2\). As shown in Fig. 1(b), we model this scenario by considering SCs deployed within a 2-D (two-dimensional) distance $d$ from the UEs, and an angle $\psi$ measured from the straight segment that links UEs and their closest mMIMO-BS. $\psi$ is chosen uniformly at random in $[-\pi/2, \pi/2]$. It is worth noting that even when the 2-D distance $d = 0$, UEs and SCs are still separated in space because the antennas are positioned at different heights, as specified in (a). In addition, to limit the effect of the inter-cell interference, we replace the Patch antenna at the SC with a more directive Yagi antenna, pointing downwards to the ground (as shown by the green radiation cone in Fig. 1(b)), and therefore only illuminating the closest UEs.

(c) **DA architecture:** There are no self-backhauled SCs deployed, and the mMIMO-BSs are solely dedicated to directly serve the UEs, as illustrated in Fig. 1(c).

1 We remark that a given SC deployed in the $i$-th sector might be connected to another mMIMO-BS $i'$ as it provides a higher RSRP level than the mMIMO-BS $i$.

2 We assume the possibility to realize this specific network deployment, for example by means of drone-BSs, where the drone-BSs can reposition themselves following the locations of UEs as suggested in [23]. Although mentioned, the drone-BSs use-case is not the focus of this paper, and it is left for future investigation.
B. Frame Structure

We consider a time-division duplexing (TDD) system, where the time-slot duration $T$ is used as a single scheduling unit in the time domain. As shown in Fig. 2(a), we partition the access and backhauling resources through the parameter $\alpha \in [0, 1]$. Therefore, the fraction $\alpha$ of time-slots is allocated to the backhaul links, and the fraction $1 - \alpha$ of time-slots is allocated to the access links. In the frequency domain, we divide the system bandwidth $BW$ into $Q_i$ RBs, and we allocate all the RBs to the backhaul links or the access links. We make the following assumptions in considering the partition of backhaul and access time-slots among the SCs and UEs:

- During the backhaul time-slots, all the associated SCs are served by the mMIMO-BS $i$ with $M$ antennas, and we use the same value of $\alpha$ for all the SCs. The mMIMO-BSs precode the backhaul signals towards the single-antenna SCs, which are spatially multiplexed in the same time-frequency resources. We consider to have $M$ DoF for a mMIMO-BS with $M$ antennas [24], and we assume $M > \mathbb{E}[L_i]$ by dimensioning the number of antennas $M$ during the mMIMO-BS deployment planning. In the case that $L_i$ may exceed the number of antennas $M$, a scheduler ensures to serve only $M$ SCs randomly picked in the set $\mathcal{L}_i$. We also observed in our experiments that $\Pr[L_i > M]$, i.e., the probability that the RV $L_i$ exceeds the number of DoF $M$, decays rapidly for large $M$, and is equal to zero for the set of system parameters adopted in our simulations.

- During the access time-slots, the SCs schedule their connected UEs by using a Round Robin (RR) mechanism, which equally distributes the available $Q_i$ RBs among its UEs.

Fig. 2(b) shows the frame structure used for the DA setup, where all the time-slots are allocated to the access links. In each time-slot, the mMIMO BSs precode the access signals, and the UEs are spatially multiplexing the entire system bandwidth. We assume $M > \mathbb{E}[K_i]$ due to the system design. When $K_i > M$, a scheduler serves only $M$ UEs randomly picked in the set $\mathcal{K}_i$. The same observation made above for the backhaul applies with $K_i$ in place of $L_i$.

Figs. 2(a) and 2(b) also show the fraction $\tau$ of the time-slots dedicated for the transmission of the uplink (UL) pilot sequences, used for the CSI acquisition. Details about the CSI acquisition procedure will be discussed in subsection II-D.

C. Channel Model

We define as $\mathbf{h}_{il} = [h_{il1}, \ldots, h_{ilm}]^T \in \mathbb{C}^M$ the propagation channel between the $l$-th single-antenna receiver (SC in the mMIMO s-BH network and UE in the mMIMO DA) and the $M$ antennas of the $i$-th mMIMO-BS. The composite channel matrix between the $i$-th mMIMO-BS and the receivers in the $i'$-th cell is represented by $\mathbf{H}_{i'i'} = [\mathbf{h}_{i'1}, \ldots, \mathbf{h}_{i'M}] \in \mathbb{C}^{M \times L_i'}$, where we omit the subscript $q$ indicating the $q$-th RB of the channel matrix for notational convenience. Moreover, for the mMIMO sBH architecture, we define the single-input single-output (SISO) channel between the $l$-th SC and the $k$-th UE in the $q$-th RB as $g_{lkq} \in \mathbb{C}$.

The channel coefficients $h_{ilm} = \sqrt{\beta_{il}} \tilde{h}_{ilm}$ and $g_{lkq} = \sqrt{\beta_{lkq}} \tilde{g}_{lkq}$ account for both the effects of the large-scale fading and the small-scale fading components:

- We model the large-scale fading components $\beta_{il}, \beta_{lk} \in \mathbb{R}^+$ according to the 3GPP Case 1 Relay scenario [18]. For a given link, the models decide whether the channel propagation conditions are LoS or NLoS, by considering a distance-dependent LoS probability function, and use log-normal distributed shadowing with different values of standard deviation. Because of its slow-varying characteristic, the large-scale fading does not change rapidly with time, and it can be assumed constant over the observation time-scale of the network.

- We model the small-scale fading components $\tilde{h}_{ilm}, \tilde{g}_{lkq} \in \mathbb{C}$, which result from multi-path, as Rician fast-fading, according to the 3GPP spatial channel model for MIMO simulations [25], assuming a $K$-factor dependent on the distance between transmitter and receiver.

Throughout the paper we assume a block-fading channel model, where the channel vectors of the access paths, i.e., the channel between each antenna at the mMIMO-BS and the UE, and the channel between the SC and the UE, remain constant for a frequency-time block corresponding to one time-slot $T$, and one RB [26]–[28]. In line with the LTE numerology [29], we consider the duration of the time-slot equal to 1 ms, and the bandwidth of the RB equal to 180 KHz. Differently, we consider that the channel vectors of the backhaul path, i.e. the channel between each antenna at the mMIMO-BS and the SC, is constant for a period $T_{BH} \gg T$ due to the static position of the SCs. Moreover, we adopt a horizontal uniform linear array (ULA) for the mMIMO-BS with equally spaced antennas and Jakes correlation model between antenna pairs [30].
D. Massive MIMO CSI Acquisition

To calculate the DL precoder of the mMIMO-BS, we consider that the channel is estimated through UL pilot sequences, assuming UL/DL channel reciprocity [3]. We also consider that the SCs or UEs associated to the same mMIMO-BS have orthogonal pilot sequences, and define the pilot codebook with the matrix \( \Phi_i = [\phi_{i1}, \ldots, \phi_{iL_i}]^T \in \mathbb{C}^{L_i \times B} \), which satisfies \( \Phi_i^H \Phi_i = I_{L_i} \). Here, the \( i \)-th sequence is given by \( \phi_{il} = [\phi_{i1l}, \ldots, \phi_{iBil}]^T \in \mathbb{C}^B \), and \( B \) denotes the pilot codebook length. Note that \( L_i \leq B \), i.e. the number of SCs trained by the mMIMO-BSs in the backhaul time-slots is limited by the maximum number of orthogonal pilot sequences. Similarly, \( K_i \leq B \), i.e. the number of UEs trained by the mMIMO-BSs in the access time-slots is limited by the maximum number of orthogonal pilot sequences. In case \( L_i > B \) or \( K_i > B \), the scheduler assigns the pilot sequences to only \( B \) SCs or UEs, randomly selected in the sets \( L_i \) and \( K_i \).

The matrix \( Y_i \in \mathbb{C}^{M_i \times B} \) of pilot sequences received at the \( i \)-th mMIMO-BS can be expressed as [31]

\[
Y_i = \sqrt{P_{il}} \sum_{i' \in I} H_{i,i'} \Phi_{i'}^H + N_i, \tag{1}
\]

where \( P_{il} \) is the power used for UL pilot transmission by the \( i \)-th device, located in the \( i \)-th sector, and \( N_i \in \mathbb{C}^{M_i \times B} \) represents the additive noise matrix, whose entries are modeled as independent and identically distributed complex Gaussian RVs with variance \( \sigma^2 \).

Let us denote by \( H_i = [h_{i1}, \ldots, h_{iL_i}] \in \mathbb{C}^{M_i \times L_i} \) the channel between the \( i \)-th mMIMO-BS and the associated devices. During the UL training phase, the mMIMO-BS obtains an estimate of \( H_i \) by correlating the received signal with a known pilot matrix \( \Phi_i \). Let us define \( \mathcal{P} \subseteq I \) as the subset of sectors, whose devices share identical pilot sequences with the devices served by the \( i \)-th mMIMO-BS. In line with other studies [31], [32], we adopt the least-squares (LS) channel estimation since we are considering a practical system, where the mMIMO-BS does not have perfect knowledge of the cross-cell channel statistics. The resulting LS channel estimation can be expressed as [31]

\[
\hat{H}_i = \frac{1}{\sqrt{P_{il}}} Y_i \Phi_i^H = H_i + \sum_{i' \in \mathcal{P}} H_{i,i'} + \frac{1}{\sqrt{P_{il}}} N_i \Phi_i^H + N_i. \tag{2}
\]

The first, second and third terms on the right-hand side of (2) represent the estimated channel, a residual pilot contamination component and the noise after the pilot sequence correlation, respectively. The use of the same set of orthogonal pilot sequences among different sectors leads to the well-known pilot contamination problem, which can severely degrade the performance of mMIMO systems [3], [33].

Similar to [33], we consider that each uplink training symbol \( \tau \) contains up to 16 orthogonal pilot sequences, and we consider its duration equal to one Orthogonal Frequency Division Multiplex (OFDM) symbol period of the Long Term Evolution (LTE) standard [29]. The pilot sequences are assigned to the SCs and to the UEs, in the case of mMIMO for backhaul and mMIMO for access, respectively.

In the mMIMO s-BH architecture, due to the static position of the SCs, the backhaul channel vectors remain constant for a longer time than the access channel vectors, and there is no need to update the CSI every time-slot. This allows to multiplex in time, over separate backhaul time-slots, the pilots transmitted by the SCs in different sectors. Therefore, we assume that no pilot contamination is present in the system, and we account for a pilot overhead \( \tau \) over \( T \), equivalent in our system model to 7% of the time-slot duration.

In the mMIMO DA architecture, we use two pilot allocation schemes [33]:

- In pilot reuse 1 (R1) the same set of pilot sequences is reused in all the sectors of the network, thus introducing strong pilot contamination. The pilot overhead to account for is \( \tau \) over \( T \), equivalent in our system model to 7% of the time-slot duration.
- In pilot reuse 3 (R3) we consider coordination between the same cell sites, avoiding using the same pilot sequence from the adjacent sectors. This reduces the effect of the pilot contamination at the expense of increasing the pilot overhead, which is \( 3 \times \tau \) over \( T \), equivalent in our system model to 21% of the time-slot duration.

In Table I, we summarize the different pilot allocation schemes used in mMIMO s-BH and mMIMO DA architectures.

| Architecture | Backhaul links | Access links |
|--------------|---------------|-------------|
| mMIMO s-BH   | pilot reuse for all the sectors (No pilot contamination) | No pilot allocation |
| mMIMO DA     | –             | • pilot reuse 1 (R1) • pilot reuse 3 (R3) |

III. DOWNLINK SINR AND USER RATE

In this section, we present the formulation for the two-hop DL data-rate in the s-BH network, which comprises the formulation for the mMIMO backhaul and the SC access SINRs and data-rates. Moreover, we include the conventional formulation for the data-rates in mMIMO DA network.

A. Massive MIMO Backhaul Transmission

The \( i \)-th mMIMO-BS uses the precoding matrix \( W_i = [w_{i1}, \ldots, w_{iL_i}] \in \mathbb{C}^{M_i \times L_i} \) to serve its connected SCs during the backhaul time-slot. In this paper, we consider that \( W_i \) is computed based on the zero-forcing (ZF) criterion as [34]

\[
W_i = \hat{H}_i^H \left( \hat{H}_i^H \hat{H}_i \right)^{-1} D_i. \tag{3}
\]

As shown in [35], among the linear methods for the mMIMO precoder, ZF and minimum mean square error (MMSE) perform better than Maximum Ratio Combining (MRC). However, ZF provides a better balance between performance and complexity than MMSE.
Here, the diagonal matrix $D_{i} = \text{diag}(\rho_{i1}, \rho_{i2}, \ldots, \rho_{iL_{i}})$ is chosen to equally distribute the total DL power $P_{i}^{\text{dl}}$ among the $L_{i}$ receivers. In the previous expression, $\rho_{il}$ represents the power allocated to the $l$-th receiver located in the $i$-th sector, and $\text{Tr}_{l}[D_{i}] = P_{i}^{\text{dl}}$, where $\text{Tr}_{l}[D_{i}]$ is the trace of matrix $D_{i}$.

Under the assumption that each SC has perfect CSI available, the DL SINR of the $l$-th slot transmitted by the $i$-th mMIMO-BS can be expressed as

$$\text{SINR}_{il}^{\text{M}} = \frac{\rho_{il} |h_{il}^{T}w_{il}|^2}{\sum_{j \neq l} \rho_{ij} |h_{ij}^{T}w_{il}|^2 + \sum_{j' \in E, j' \neq i} \rho_{ij'} |h_{ij'}^{T}w_{il}|^2 + \sigma_{n}^2}.$$  \hspace{1cm} (4)

The numerator of (4) contains the power of the signal intended for the $l$-th receiver, and the denominator includes the intra-cell interference from the serving $i$-th mMIMO-BS, the inter-cell interference from other mMIMO-BSs, and the power of the thermal noise at the SC receiver $\sigma_{n}^2$.

The corresponding DL backhauling rate at the $l$-th SC receiver can therefore be expressed as

$$R_{il}^{\text{dl}} = \alpha \left(1 - \frac{T}{T}ight) BW \log_2 \left(1 + \text{SINR}_{il}^{\text{M}}\right).$$  \hspace{1cm} (5)

where $\alpha$, as indicated before, represents the fraction of time-slots allocated to the backhaul links.

B. Small Cell Access Transmission

We recall from the channel model that $g_{lkq}$ denotes the SISO channel between the $l$-th SC and the $k$-th UE corresponding to the $q$-th RB. The DL SINR of the $k$-th UE served by the $l$-th SC in RB $q$ can be expressed as

$$\text{SINR}_{lkq}^{\text{M}} = \frac{P_{l}^{\text{dl}} |g_{lkq}|^2}{\sum_{i \in I} \sum_{l' \neq l} P_{l'}^{\text{dl}} |g_{lrkq}|^2 + \sigma_{n}^2},$$  \hspace{1cm} (6)

where $P_{l}^{\text{dl}}$ and $P_{l'}^{\text{dl}}$ are the transmit powers on the RB of the $l$-th and $l'$-th SCs, respectively, and $\sigma_{n}^2$ denotes the thermal noise power at the UE receiver.

The corresponding DL access rate for UE $k$ served by SC $l$ can be therefore expressed as

$$R_{lk}^{\text{Da}} = (1 - \alpha) \frac{BW}{Q_{l}} \sum_{q=1}^{Q_{l}} x_{q}^{k} \log_2 \left(1 + \text{SINR}_{lkq}^{\text{M}}\right),$$  \hspace{1cm} (7)

where $x_{q}^{k} = 1$ if the $q$-th RB is assigned to the $k$-th UE, and $x_{q}^{k} = 0$ otherwise.

The potential aggregated DL access rate provided by the $l$-th SC is $R_{lk}^{\text{DA}} = \sum_{k=1}^{K_{l}} R_{lk}^{\text{Da}}$. However, the actual aggregated DL access rate provided by the $l$-th SC cannot be larger than the backhaul DL rate, which entails that $R_{il}^{\text{Da}} \leq R_{il}^{\text{dl}}$, $\forall l \in L_{i}$, and $\forall i \in I$. In this paper, we assume that the backhaul capacity is equally divided between the $K_{l}$ UEs served by the $l$-th SC.

The assumption of equally distributed backhaul capacity among the UEs might become a drawback for the end-to-end rates, when UEs served by the same SC have different rate requirements in the access links, and in this case, the partition of the backhaul resources among the UEs could be designed according to their demands. This access-based partition of the backhaul resources among the UEs is not the focus of this paper, and its study in the context of s-BH architecture is left for future work.

Therefore, the resulting end-to-end access rate for the $k$-th UE can be expressed as

$$R_{ik} = \min \left(\frac{R_{il}^{\text{Da}}}{R_{il}^{\text{dl}}}, \frac{R_{il}^{\text{DA}}}{R_{il}^{\text{dl}}}\right),$$  \hspace{1cm} (8)

where the minimum is computed between the backhaul rate obtained from (5), divided by the number of UEs served by the $l$-th SC, and the access rate obtained from (7).

C. Massive MIMO Direct Access Transmission

In contrast to s-BH setups, mMIMO systems providing DA dedicate all their time resources to multiplex DL data streams to the UEs. Thus, the DL access rate of the $k$-th UE served by the $i$-th mMIMO-BS can be expressed as

$$R_{ik}^{\text{DA}} = \left(1 - \frac{T}{T}\right) BW \log_2 \left(1 + \text{SINR}_{ik}^{\text{DA}}\right),$$  \hspace{1cm} (9)

where the estimated channel matrix $\tilde{H}_{i} = [\tilde{h}_{i1} \cdots \tilde{h}_{iK_{i}}] \in \mathbb{C}^{M \times K_{i}}$ between the $i$-th mMIMO-BS and its connected UEs is plugged into (4), to subsequently derive the DL SINR in (4), assuming that each UE has perfect CSI available, and the access rate in (9).

IV. ANALYTICAL SIR AND AVERAGE BACKHAUL AND ACCESS RATES

In the following, we present a tractable formulation to model the mMIMO s-BH network, which approximates the backhaul and access SINR and data-rate expressions.

We recall from Section III that the mMIMO-BS locations are distributed on a hexagonal grid with density $\lambda_{a} = 3(2\pi R^{2})^{-1}$, where $R = \sqrt{\frac{d_{1SD}}{2\pi}}$ is the outer sector radius of the hexagonal site, and $d_{1SD}$ is the inter-site distance between two cell sites. For the random deployment, we consider that the locations of SCs and UEs are drawn from two independent SPPPs $\Omega_{a}$ and $\Omega_{u}$ with densities $\lambda_{a}$ and $\lambda_{u}$, respectively. For the ad-hoc deployment, we consider that the SC locations are coupled to the UE locations drawn from a SPPP $\Omega_{a}$, and we position the SCs at a fixed distance $d$ with respect to the UE locations. The mean numbers of SCs and UEs in the finite area of the sector are obtained as $\mu_{u} = \lambda_{u}/\lambda_{a}$ and $\mu_{a} = \lambda_{u}/\lambda_{a}$, respectively [36]. Moreover, when considering a dense deployment of SCs, there is a high probability that there are SCs unloaded, i.e. without UEs associated to them [37]. We capture this effect in the analysis by considering an activation probability for the SC, which can be approximated as [38]

$$p_{a} \approx 1 - \left(1 + \frac{\lambda_{u}}{3.5\lambda_{b}}\right)^{-3.5},$$  \hspace{1cm} (10)

where the expression $\left(1 + \frac{\lambda_{u}}{3.5\lambda_{b}}\right)^{-3.5}$ approximates the average void probability for a typical Voronoi cell [38]. Consequently, the locations of the active SCs can be approximated as SPPP $\tilde{\Omega}_{b}$ with density $\tilde{\lambda}_{b} = p_{a}\lambda_{b}$ derived by thinning the SC process $\Omega_{b}$ [38]. Differently, in the ad-hoc deployment we assume $p_{a} = 1$, meaning that each SC is active and serves the nearby UE. The mean number of active SCs in a sector is obtained as $\tilde{\mu}_{b} = \lambda_{b}/\lambda_{a}$.
In the following expressions, SIR is used to approximate the SINR, since in the sub-6 GHz bands, with a system design which assures signal coverage, the system operates in interference-limited conditions, where the power of received interference dominates the denominator of the SINR.

A. Average Rates of Massive MIMO Backhaul Transmission

We now provide an analytical model for evaluating the average data-rate of the backhaul links, given the SIR of a typical SC and the spatial distribution of SCs in the sector. Inspired by [20], we treat the SCs as UEs, and we extend the framework proposed in [19] to model the SIR by considering: i) the effects of antenna directionality and sectorization, captured with the horizontal and vertical antenna patterns, and by modeling the co-site interference component; ii) the effect of the beamforming gain due to the mMIMO precoding as proposed in [14]. We make the following assumptions in our analytical backhaul model:

- For simplicity, the backhaul channel is statistically modeled by considering only the large-scale fading component, excluding shadowing statistics.
- We assume LoS propagation channel conditions from the serving mMIMO-BS to the SC (and from the co-site interfering mMIMO-BSs to the SC), to reflect the characteristics of the backhaul link, which tends to have dominant LoS conditions between SCs and the antennas of the nearest mMIMO-BSs [18]. On the other hand, we assume all the mMIMO-BSs from the surrounding interfering sites to be in NLoS.
- We assume that the co-channel interference from the mMIMO-BS to other served SCs can be reasonably neglected since we adopt the ZF precoder. For consistency with the assumptions made in Section II-D and in Section III-A, we consider that there is no pilot contamination in the backhaul links, and we consider that each SC has perfect CSI available.

For the analysis that follows, \( r \) and \( \theta \) denote two independent RVs, which define the distance and the angle from the SC to the serving mMIMO-BS. Note that \( r \) and \( \theta \) are distributed with uniform probability density functions (pdfs) \( f_r(r) \) and \( f_\theta(\theta) \) in the interval \([r_{\text{min}}, \frac{d_{\text{ISD}}}{2}]\) and \([-\pi/3, +\pi/3]\), respectively, where the distance \( r_{\text{min}} \) denotes the minimum distance between the mMIMO-BS and the SC. By convention, \( \theta = 0 \) indicates the boresight direction in the first sector, and such sector is denoted as \( s = 1 \) for each hexagonal cell formed by \( S \) sectors.

The SIR associated to the mMIMO-BS is modeled as

\[
\text{SIR}^B(r, \theta) \approx \frac{(M - \hat{\mu}_b + 1)/\hat{\mu}_b}{P_i^i G_a G_V(r) G_{H,1}(\theta) \beta^L(r)} I_1(r, \theta) + I_2(r),
\]

where the multiplying factor \((M - \hat{\mu}_b + 1)/\hat{\mu}_b\) represents the beamforming gain from mMIMO precoding \( G_a \), \( G_V(r) \) and \( G_{H,1}(\theta) \) are the antenna gains of the single mMIMO-BS element, the vertical \( V \), and the horizontal \( H \) antenna patterns, respectively. \( \beta^L(r) = A^L r^{-\eta^L} \) is the pathloss between mMIMO-BS and SC, where \( A^L \) and \( \eta^L \) indicate the frequency dependent pathloss factor and the pathloss exponent for the backhaul link in LoS condition, respectively, and \( I_1(r, \theta) \) and \( I_2(r) \) are the co-site and inter-site interference components, respectively. The vertical antenna pattern is defined as [18]

\[
G_V(r)_{|\text{DBI}} = -\min \left( 12 \left( \frac{\tan(\delta_a / \sqrt{r^2 - \delta_a^2} - \frac{\pi}{2} \delta_H)}{\frac{\pi}{2} \delta_H} \right)^2, F_v \right),
\]

where \( \delta_a \) is the difference in antenna heights between the mMIMO-BS and the SC, \( \delta_H \) is the mechanical downtilt, \( \delta_H \) is the half-power vertical beamwidth, and \( F_v \) is the vertical front-back ratio. Similarly, the horizontal antenna pattern is defined as [18]

\[
G_{H,s}(\theta)_{|\text{DBI}} = -\min \left( 12 \left( \frac{\theta - (s-1)2\pi/3}{\theta_H} \right)^2, F_h \right),
\]

where \( \theta_H \) is the half-power horizontal beamwidth, and \( F_h \) is the horizontal front-back ratio.

In (11), \( I_1(r, \theta) \) is given by

\[
I_1(r, \theta) = P_i^i d_b G_a G_V(r) \sum_{s=2}^S G_{H,s}(\theta) \beta^L(r),
\]

and \( I_2(r) \) is approximated as [19]

\[
I_2(r) \approx \frac{2\pi \lambda_s P_i^i d_b G_a G_V(2R_c - r) G_H A^{NL}}{\eta^{NL} - 2} \times \left( (2R_c - r)^{2-\eta^{NL}} - (R_b - r)^{2-\eta^{NL}} \right),
\]

where \( R_b = \frac{3}{2}d_{\text{ISD}} \) denotes the network boundary, \( R_c = \frac{d_{\text{ISD}}}{2} \) is the inner sector radius, \( G_H = \int_\theta^{2\pi} \sum_{s=1}^S G_{H,s}(\theta) d\theta \) is the average horizontal antenna gain with respect to \( \theta \), and \( \eta^{NL} \) and \( A^{NL} \) are the pathloss exponent and the frequency dependent pathloss factor for the backhaul link in NLoS condition, respectively. Eq. (15) approximates the integration area used to calculate the inter-cell interference from other macro-BSs assuming a ring formed by two concentric discs centered at the origin and with radii of \( 2R_c \) and \( R_b \). We refer to [19] for the validation of this model.

Finally, the average SC data-rate for backhaul transmission can be expressed as

\[
\overline{R^B} = \alpha BW \times \int_{-\frac{\pi}{3}}^{\frac{\pi}{3}} \int_{r_{\text{min}}}^{R_c} \log_2 (1 + \text{SIR}^B(r, \theta)) f_r(r, \theta) d\theta dr.
\]

Therefore, the results for the average SC data-rate can be computed by numerical integration of (16).

\footnote{We observed that for the value of \( r_{\text{min}} \) that we use in our simulations, the realizations of the (modified) spatial process could be statistically considered as if they were derived from a homogeneous SPPP, since we tested the spatial randomness with the Ripley’s K-function [29], and we verified that the theoretical values of the SPPP are contained within the region defined by the lower and upper envelopes given by the empirical results.}

\footnote{Only the active SCs are spatially multiplexed in the backhaul time-slots, since those inactive are not required to backhaul the UEs data.}
B. Average Rates of Small Cell Access Transmission

Inspired by the stochastic geometry analysis presented in [40], we now provide an analytical model for evaluating the access SIR of a typical UE at the origin, and its access average DL data-rate. Similarly to [41], [42], we consider the impact of the LoS and NLoS pathloss characteristics to model the SIR. We use the same LoS probability function as in [42], however, we consider for the inter-cell interference computation the density of the active SCs, i.e. those with UEs associated.

We make the following assumptions in our analytical access model:

- We assume that each UE connects to the nearest SC. For the random deployment, since \( \Omega_b \) and \( \Omega_a \) are two independent SPPPs, we assume that \( x \), i.e. the distance between UE and the serving SC is a RV Rayleigh distributed with pdf \( f_X(x) = 2\pi\lambda_b x \exp(-\lambda_b\pi x^2)/\exp(-\lambda_b\pi \delta_b^2) \), where \( \delta_b \) denotes the difference between the SC and UE heights [40]. Differently, for the ad-hoc deployment, we assume \( x = \sqrt{d^2 + \delta_b^2} \).
- The propagation channels are represented with a combination of distance-dependent pathloss and multi-path fading, distributed as Rayleigh with an exponential power distribution \( |g|^2 \sim \exp(1) \).
- We adopt a probabilistic LoS channel model for the inter-cell interference, with a LoS probability expressed as [42]

  \[
  \Pr[\text{LoS}] = \exp(- (x/D)^2), \quad (17)
  \]

  where the parameter \( D \) is set to approximate the LoS probability of the SC-UE 3GPP model [42].

  The SIR of a typical UE at the origin associated to the SC is modeled as

  \[
  \text{SIR}^A(x) \approx \frac{P_A^{10}G_b |g|^2 \beta^2 (x)}{I_{agg}}, \quad (18)
  \]

  where \( G_b \) is the SC antenna gain, \( |g|^2 \) is the multi-path channel gain, and \( I_{agg} \) is the aggregated inter-cell interference.

  We now use [18] to provide an expression for the rate coverage probability, which defines the probability that the UE rate is higher than a minimum target \( R_{th} \). This probability can be expressed as \( \Pr [\text{SIR}^A(x) > \gamma_a] \), where \( \gamma_a = 2^{R_{th}/\mu_2} - 1 \) depends on \( R_{th} \), and we approximate the number of UEs associated with the SC serving the typical UE at the origin with the corresponding mean \( \mu_2 \) given by \( 1+1.28\lambda_a/\lambda_b \) [43]. We later show in Fig. 5(b) that the mean of the distribution of the number of UEs for a typical SC, given by \( \lambda_a/\lambda_b \) [44], matches the numerical results.

The expressions used to evaluate the rate coverage probability are included in Appendix A. Thus, the average UE data-rate for access transmission can be expressed as

\[
R^A = (1 - \alpha) BW \times \int_0^{\infty} \int_{\delta_b}^{\infty} \Pr [\text{SIR}^A(x) > \gamma_a] f_X(x) dx d\gamma_a. \quad (19)
\]

Therefore, the results for the average UE data-rate can be computed by numerical integration of (19).

In the next section, we will use this model to complement the insights given by the 3GPP-based system-level simulations.

V. SIMULATIONS AND NUMERICAL RESULTS

In this section, we evaluate the performance of mMIMO s-BH and DA networks using 3GPP-based system-level simulations and mathematical analysis.

In the mMIMO s-BH network, the different characteristics of the backhaul and access radio links are modeled considering the methodology described in [18] for the 3GPP Case 1 Relay scenario. As described in [18] Tab. A.2.1.1.2-3, we adopt the LoS and NLoS pathloss exponents \( \eta^L = 2.35 \) and \( \eta^NL = 3.63 \) for the backhaul links, and we consider \( \eta^L = 2.09 \) and \( \eta^NL = 3.75 \) for the access links. For each link, we use the corresponding LoS probability function proposed in [18] Tab. A.2.1.1.2-3. To simulate the backhaul links, we account...
for the SC site planning correction factor, which affects the pathloss and the LoS probability as indicated in [18, Tab. A.2.1.1.4-2]. To simulate the access links, we assume cross-correlated shadowing, with correlation coefficient $\rho = 0.5$ at the UE location with respect to the different SCs [18, Tab. A.2.1.1.2-3]. In the simulations, we consider a Rician fading model, and we characterize the Rician $K$-factor with the model: $K[dB] = 13 - 0.03r$ in dB, where $r$ is the distance between transmitter and receiver in meters [25].

In the 3GPP-based system-level simulations, the channel gains (composed by pathloss, shadowing and multi-path fading) are generated for all useful and interfering radio links between each SC and the UEs, as well as between each mMIMO-BS and all SCs. We collect statistics for different network realizations, each with independent deployments of UEs and SCs. Subsequently, we measure the performance in terms of the cumulative distribution functions (CDFs) of the backhaul SC rates in (5), of the access UE rates in (7), and of the end-to-end UE rates in (8).

To compare mMIMO s-BH against mMIMO DA architectures, we also simulate the links between mMIMO-BSs and UEs, and compute the resulting rates in (9). In the mMIMO DA network, we adopt the LoS probability function and the corresponding exponents $\eta^{L} = 2.42$ and $\eta^{NL} = 4.28$, as indicated in [18, Tab. A.2.1.1.2-3].

Table II contains the relevant parameters used to conduct the simulation campaign.

**A. Massive MIMO s-BH: Random vs. Ad-hoc Small Cell Deployments**

In this subsection, we analyze the 3GPP-based simulation results for two SC topologies described in Sec. II-A, namely the ad-hoc and random SC deployments. In both cases, $\mu_u = 16$ UEs are deployed per sector on average and scheduled in access time-slots by their serving SCs. We evaluate the impact of densification by considering $\mu_b = \{4, 8, 16\}$ SCs per sector on average for the case of random SC deployments. In the ad-hoc deployment, we consider $\mu_b = 16$ SCs per sector on average, and different values of the 2-D distance $d$ from the UE to the SC. The resource partition $\alpha$ is set to 0.5, to distribute between backhaul and access the available resources equally.

As a first step, we compare the LoS probability of the backhaul and access links. The group of results in the left part of Fig. 3 shows the probability of a given SC to be in LoS with respect to the server mMIMO-BS with different densities of SCs and deployments. As expected, the percentages of backhaul links in LoS are almost the same in both the random and ad hoc deployments, since in the first case the SCs are randomly distributed with respect to mMIMO-BSs, while in the last approach the SCs are positioned in the vicinity of the UEs, which are randomly distributed with respect to mMIMO-BSs. Moreover, we can also see that the LoS probability of a UE in the s-BH architecture increases as the density of SCs increase, reaching 100% probability of LoS channel condition in the ad-hoc deployment, as shown by the results in the right part in Fig. 3. Overall, the backhaul link mainly limits the joint backhaul-access probability of LoS-LoS conditions.

As a second step, Figs. 4(a), 5(a) and 6 analyze the data-rate performance of the backhaul and access transmission by first considering the two links separately, and then their combined effect on the end-to-end UE rate. The following considerations can be made:

- **Backhaul link performance:** Fig. 4(a) illustrates the CDFs of the backhaul data-rates. These results show how increasing the number of SCs randomly deployed, and especially with the ad-hoc deployment, the backhaul data-rate received by each SC decreases. This is due to the reduction of the multiplying factor $[(M - \tilde{\mu}_b + 1)P^{dl}] / \tilde{\mu}_b$ in (11), and the split of the transmit power among the active backhaul streams. It is worth to note that only the SCs with associated UEs are active (i.e. transmitting to the UEs), and are served via multiple backhaul links. Thus, looking at Fig. 4(b) which show the average number of SCs served by the mMIMO-BS when applying the random and ad-hoc SCs deployment strategies, we can better explain the results presented in Fig 4(a). In fact, while with the ad-hoc deployment almost all the 16 SCs are always active, with 16 randomly deployed SCs only 10 of them are active in average, as a result of the UEs association procedure. As depicted in Fig. 4(b) the analytical approximation in (10) of the SC activation probability matches the numerical results obtained by simulations.

- **Access link performance:** Fig. 5(a) shows the results for the access data-rate. As a general conclusion we can see that adding more randomly deployed SCs in the sector does not introduce a significant gain, while opportunistically deploying one SC closer to each UE is quite beneficial. In the following, we discuss the details of the different factors playing a key role in these results. On the one hand, when densifying the network, the carrier signal benefits from having SCs that are more likely in close vicinity to the served UE, even if a random SC deployment does not always guarantee this vicinity. Also, each SC has to serve a progressively reduced number of UEs in the access links (as indicated in Fig. 5(b)), and accordingly in the backhaul links, which means having
more RBs available to allocate to each UE over different links. On the other hand, adding more SCs increases the probability of having a larger number of LoS interferers at the UE side. In the random deployment, the power of the interference links grows faster than the carrier signal power due to NLoS to LoS transition of the interference links [41]. In the ad-hoc case, the same interference effect takes place. However, by decreasing the distance $d$ from UE to SC in such a way as to ensure that the position of the SC is always close to the served UE, the power of the carrier signal increases faster than the interference power, and thus the hit in the SINR is not as significant. As a result, only a very dense deployment of random SCs could provide the same data-rate as in the case of the ad-hoc deployment. In subsection V-D, we will discuss the asymptotic behavior when increasing the density of SCs, and quantify the number of required randomly deployed SCs to achieve this condition.

From Fig. 5(a) we can observe how equipping the SC with a more directive antenna (i.e. Yagi) with respect to the Patch antenna further boosts the access link capacity to achieve 75 Mbps per UE at the median value. Two complementary effects cause this performance enhancement: i) the signal improvements provided by the higher antenna gain of the directive Yagi, and ii) the reduced interference created towards neighboring UEs served by other SCs.

- **End-to-end overall performance:** Fig. 6 shows the end-to-end results given by the combination of the two-hop, backhaul and access, performance previously depicted in Figs. 4(a) and 5(a). Overall, the end-to-end data-rates of the random deployment are more limited by the access links than by the backhaul links, as shown by comparing...
the results in Figs. 6 to the one in 5(a). On the contrary, the end-to-end data-rates of the ad-hoc deployment outperform the one of the random deployment, but are severely penalized by the backhaul links, as shown by comparing the results in Figs. 6 to the one in 4(a).

Thus, the reduced backhaul capacity of the mMIMO sBH ad-hoc deployment does not fully allow to exploit the potentially high data-rate achieved in the access. This behavior suggests the need to optimize the splitting of resources between the two links. Indeed, a particularly important improvement in end-to-end rates would be achieved through the allocation of more resources to backhaul links. This is analyzed in the following section.

B. Massive MIMO s-BH: Access and Backhaul Resource Allocation

In Fig. 7, we vary $\alpha$ in the range $0 \leq \alpha \leq 1$, and analyze the behavior of UEs rate at the 5-th and 50-th percentiles of the CDF. The configurations $\alpha = 0$ and $\alpha = 1$ entail that all the time-slots are allocated to the access and the backhaul, respectively. Therefore, the UE rates for these two values are equal to 0, since no resources are left for the other link. Moreover, the configuration $\alpha^*$ represents the value of $\alpha$ that maximizes the UE rate. For instance, with $d = 0$ and Yagi antennas at the SCs, $\alpha^*$ is equal to 0.85 when looking at the 5-th percentile. Fig. 7 brings the following insights:

- By comparing the results between Fig. 7(a) and Fig. 7(b), it is important to note that the optimal $\alpha$ changes from 0.85 to 0.75. A tradeoff exists between 5-th and 50-th percentile performance, and they cannot be optimized simultaneously. Assuming that the network uses $\alpha = 0.85$, which is the optimal value for cell-edge UEs (5-th percentile of the CDF), the median UEs (50-th percentile of the CDF) can achieve an end-to-end rate of 19.5 Mbps, which represents a 16% reduction with respect to the maximum end-to-end rate achievable of 23.3 Mbps with $\alpha = 0.75$.

- In Figs. 7(a) and 7(b), we show with dashed lines the results of the mMIMO DA setup. A properly designed mMIMO s-BH radio resource partitioning can improve the performance of the cell-edge UEs, and keeps the same performance for the UEs at the median of the CDF with respect to mMIMO DA architecture, as shown in Figs. 7(a) and 7(b), respectively. A more detailed comparison is further developed in the next section.

C. Massive MIMO Architectures: s-BH vs. Direct Access

First, we compare the joint probability of LoS-LoS channels in the backhaul-access legs for the s-BH network with the probability of LoS channel in the access links for the DA network. As shown in Fig. 8, there are higher joint backhaul-access LoS probabilities with respect to the access LoS probability, 47% and 25%, respectively. Those are the cases where the s-BH architecture can potentially improve the UE performance with respect to the DA architecture.
We also compare the end-to-end UE rates resulting from the 3GPP-based simulations of the mMIMO s-BH and mMIMO DA networks. As shown in Fig. 9, the mMIMO s-BH network with the ad-hoc deployment of SCs provides better performance than the mMIMO DA network with pilot reuse 1 at the bottom of the CDF, i.e. below the 50-th percentile. In fact, pilot contamination severely degrades the rates of UEs at the cell edge in the mMIMO DA setup with pilot reuse 1. On the other hand, in the s-BH network, due to the longer coherence time of the backhaul channel, $T_{BH}$, with respect to the system timeslot duration, $T$, there is no pilot contamination, and the UEs benefit from the proximity of SCs, which reduces the pathloss and improves the LoS propagation condition, as shown in Fig. 8.

However, by adopting the pilot reuse 3 in the mMIMO DA network, the pilot contamination effect reduces, and the results show that the mMIMO DA performance exceeds the one of the mMIMO s-BH with $\alpha = 0.5$, even if the pilot overhead ($\tau = 3$ OFDM symbols) is 3 times larger with respect to pilot reuse 1. The mMIMO s-BH architecture provides the same performance as the mMIMO DA for the median UEs, only when the optimal partition $\alpha = \alpha^*$ is selected. Overall, mMIMO s-BH underperforms mMIMO DA above the median of the UE rates, and provides rate improvements for cell-edge UEs that amount to $30\%$ and a tenfold gain when adopting pilot reuse 3 and reuse 1, respectively.

**D. Asymptotic Data-rate Analysis**

Figs. 10(a) and 10(b) show the convergence behavior of the backhaul and access data-rates for the random SCs deployment with respect to the results obtained with the ad-hoc deployment of 16 SCs, positioned at fixed distance $d = 0$ with respect to the UEs. In both cases the results are obtained by numerical integration of (16) and (19). Fig. 10(a) show the convergence results of the backhaul link. The backhaul data-rate of the random distribution (solid line) converges to the one of the ad-hoc (dashed line) when the number of SCs is 100 times larger than the number of ad-hoc SCs (denoted by the marker “*”). Fig. 10(b) shows the convergence results of the access link. In this case, the data rate of the random (solid line) converges to the ad-hoc (dashed line), when the number of SCs is 1000 times larger than the number of SCs deployed in the ad-hoc case (marker “*”). As shown in subsection V-B, the ad-hoc deployment is the one that maximizes the end-to-end data-rates of the two-hop communication.

The main takeaway is summarized by the possibility for the MNOs to explore the adoption of future dynamic SCs infrastructures [45]. Indeed, instead of significantly over-
provisioning the number of SCs, it may be beneficial to dynamically reposition only the active ones, trying to guarantee the same performance that are obtained with a very dense deployment of SCs.

VI. CONCLUSION AND FUTURE WORKS

In this paper, we studied the performance results for two 5G mMIMO architectures working at frequencies below 6 GHz: mMIMO s-BH and mMIMO DA. In the mMIMO s-BH architecture, we analyzed two different configurations: random deployment of SCs in the coverage area of the serving macro cell, and ad-hoc deployment of SCs in close proximity to each UE. The first takeaway of this study is that the random SCs distribution entails deploying thousands of SCs to achieve the access link performance upper bound. On the other hand, the ad-hoc deployment benefits from the close proximity of the SCs to the UEs, and outperforms the random one for reasonable numbers of deployed SCs. However, the SC requires to know the UE position, and this is particularly complicated to realize due to mobility. The second takeaway is that an ad-hoc SCs deployment supported by mMIMO s-BH provides rate improvements for cell-edge UEs that amount to 30% and a tenfold gain as compared to mMIMO DA with pilot reuse 3 and reuse 1, respectively. This means that pilot contamination severely impacts the UEs performance at the cell-edge, and it is better to serve the UEs with the mMIMO s-BH architecture. On the other hand, mMIMO DA outperforms s-BH above the median of the UE rates, meaning that when pilot contamination is less severe, and the LoS probability of the DA links improves, is better to serve the UEs with the mMIMO DA architecture.

Future works will focus on improving the capacity of the backhaul, which is currently the main limitation in the s-BH architecture at the sub-6 GHz frequencies. At first, we aim to investigate how to improve the LoS probability in the backhaul links. This improvement can be achieved by increasing the altitude of the s-BH SCs, for example by means of drone-Bs hovering above the UE locations. Secondly, we aim to increase the bandwidth and limit the interference of the backhaul links. This can be realized by using the mmWave frequencies which can simultaneously guarantee the availability of abundant spectrum in the mmWave bands, and partially avoid the inter-cell interference issue, due to the high directionality of the transmission at those carrier frequencies.

ACKNOWLEDGMENT

The authors would like to thank Dr. Jacek Kibilda for his valuable suggestions and helpful comments.

REFERENCES

[1] A. Bonfante, L. Galati Giordano, D. Lopez-Perez, A. Garcia-Rodriguez, G. Geraci, P. Baracca, M. M. Butt, M. Dzaferagic, and N. Marchetti, “Performance of massive MIMO self-backhauling for ultra-dense small cell deployments,” in Proc. IEEE Global Commun. Conf. (Globecom), Dec. 2018, pp. 1–7.

[2] A. Gupta and R. K. Jha, “A survey of 5G network: Architecture and emerging technologies,” IEEE Access, vol. 3, pp. 1206–1232, Jul. 2015.
[3] F. Rusak, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, and T. Ufvesson, “Scaling up MIMO: Opportunities and challenges with very large arrays,” IEEE Signal Processing Mag., vol. 30, no. 1, pp. 40–60, Jan. 2013.

[4] D. López-Pérez, M. Ding, H. Claussen, and A. H. Jafari, “Towards 1 Gbps/UE in cellular systems: Understanding ultra-dense small cell deployments,” IEEE Communications Surveys Tutorials, vol. 17, no. 4, pp. 2078–2101, Fourthquarter 2015.

[5] U. Siddique, H. Tabassum, E. Hossain, and D. I. Kim, “Wireless backhauling of 5G small cells: challenges and solution approaches,” IEEE Wireless Communications, vol. 22, no. 5, pp. 22–31, Oct. 2015.

[6] N. Wang, E. Hossain, and V. K. Bhargava, “Backhauling 5G small cells: A radio resource management perspective,” IEEE Wireless Communications, vol. 22, no. 5, pp. 41–49, Oct. 2015.

[7] 3GPP TR 38.874, “Study on integrated access and backhaul,” Technical Report (TR), May 2018.

[8] M. N. Kulkarni, J. G. Andrews, and A. Ghosh, “Performance of dynamic and static TDD in self-backhauled millimeter wave cellular networks,” IEEE Trans. Wireless Commun., vol. 16, no. 10, pp. 6460–6478, Oct. 2017.

[9] S. Seifgh, M. N. Kulkarni, A. Ghosh, and J. G. Andrews, “Tractable model for rate in self-backhauled millimeter wave cellular networks,” IEEE J. Sel. Areas Commun., vol. 33, no. 10, pp. 2196–2211, Oct. 2015.

[10] R. Gupta and S. Kalyanasundaram, “Resource allocation for self-backhauled networks with half-duplex small cells,” in Proc. IEEE Int. Conf. on Comm. (ICC), May 2017, pp. 198–204.

[11] T. M. Nguyen, A. Yadav, W. Ajib, and C. Assi, “Resource allocation in two-tier wireless backhaul heterogeneous networks,” IEEE Trans. Wireless Commun., vol. 15, no. 10, pp. 6690–6704, Oct. 2016.

[12] A. Sharma, R. K. Ganti, and J. K. Milleth, “Joint backhaul-access analysis of full duplex self-backhauling heterogeneous networks,” IEEE Trans. Wireless Commun., vol. 16, no. 3, pp. 1727–1740, Mar. 2017.

[13] B. Li, D. Zhu, and P. Liang, “Small cell in-band wireless backhaul in massive MIMO systems: A cooperation of next-generation techniques,” IEEE Trans. Wireless Commun., vol. 14, no. 12, pp. 7057–7069, Dec. 2015.

[14] H. Tabassum, A. H. Sakr, and E. Hossain, “Analysis of massive MIMO-enabled downlink wireless backhauling for full-duplex small cells,” IEEE Trans. Commun., vol. 64, no. 6, pp. 2354–2369, Jun. 2016.

[15] H. H. Yang, G. Geraci, and T. Q. S. Quek, “Energy-efficient design of MIMO heterogeneous networks with wireless backhaul,” IEEE Trans. Wireless Commun., vol. 15, no. 7, pp. 4914–4927, Jul. 2016.

[16] M. Feng, S. Mao, and T. Jiang, “Joint frame design, resource allocation and user association for massive MIMO heterogeneous networks with wireless backhaul,” IEEE Trans. Wireless Commun., vol. 17, no. 3, pp. 1937–1950, Mar. 2018.

[17] Y. G. Lim, C. B. Chae, and G. Caire, “Performance analysis of massive MIMO for cell-boundary users,” IEEE Trans. Wireless Commun., vol. 14, no. 12, pp. 6827–6842, Dec. 2015.

[18] 3GPP TR 36.814, “Further advancements for E-UTRA physical layer aspects,” Technical Report (TR), Mar. 2017.

[19] J. M. Kelif, M. Couppechoux, and P. Godlewski, “A fluid model for performance analysis in cellular networks,” EURASIP Journ. Wireless Commun. Network., vol. 2010, no. 435189, Aug. 2010.

[20] M. Minelli, M. Ma, M. Couppechoux, J. M. Kelif, M. Sigelle, and P. Godlewski, “Optimal relay placement in cellular networks,” IEEE Trans. Wireless Commun., vol. 13, no. 2, pp. 998–1009, Feb. 2014.

[21] V. Chandrasekhar and J. G. Andrews, “Spectrum allocation in tiered cellular networks,” EURASIP J. Wireless Commun. Network., vol. 57, no. 10, pp. 3059–3068, Oct. 2009.

[22] ——, “Uplink capacity and interference avoidance for two-tier femtocell networks,” IEEE Trans. Wireless Commun., vol. 8, no. 7, pp. 3498–3509, Jul. 2009.

[23] C. Liu, M. Ding, C. Ma, Q. Li, Z. Lin, and Y. Liang, “Performance analysis for practical unmanned aerial vehicle networks with LoS/NLoS transmissions,” in Proc. IEEE Int. Conf. on Comm. Workshops (ICC Workshops), May 2018, pp. 1–6.

[24] H. Huh, G. Caire, H. C. Papadopoulos, and S. A. Ramprasad, “Achieving Massive MIMO spectral efficiency with a not-so-large number of antennas,” IEEE Trans. Wireless Commun., vol. 11, no. 9, pp. 3226–3239, Sep. 2012.

[25] 3GPP 25.996, “Spatial channel model for Multiple Input Multiple Output (MIMO) simulations (Release 14),” Technical Report (TR), Mar. 2017.

[26] T. L. Marzetta, E. G. Larsson, H. Yang, and H. Q. Ngo, “Fundamentals of massive MIMO,” Cambridge University Press, 2016.
In the following, we derive the expression used to evaluate the rate coverage probability. The probability that the access SIR is greater than a threshold $\gamma_a = \gamma_{\text{Rh1,\mu}} - 1$, which depends on the minimum target rate $R_{th}$, is expressed as

$$
\Pr \left[ \text{SIR}^A(x) > \gamma_a \right] = \Pr \left[ \frac{P_{\text{th}}^L G_0 |g|^2 \beta^L(x)}{I_{\text{agg}}(s)} > \gamma_a \right] = \mathcal{L}_{I_{\text{agg}}}(s) \left( \frac{\gamma_a}{P_{\text{th}}^L |G_0\beta^L(x)|} \right),
$$

where $(a)$ follows from [41] eq. (54)] neglecting the thermal noise as the propagation in sub-6 GHz bands is interference-limited. The $\mathcal{L}_{I_{\text{agg}}}(s)$ represents the Laplace transform of $I_{\text{agg}}$, and is defined according to [41], [42] as follows

$$
\mathcal{L}_{I_{\text{agg}}}(s) = \exp \left( -\pi \lambda_x \int_{x_1}^{+\infty} \frac{1}{1 + \left(s P_{\text{th}}^L G_0 |G_0\beta^L(u)|\right)^{-1}} du \right) \times \exp \left( -2\pi \lambda_x \int_{x_1}^{+\infty} \frac{1}{1 - \left(s P_{\text{th}}^L G_0 |G_0\beta^L(u)|\right)^{-1}} du \right),
$$

where $x_1 = \left( \frac{\lambda_c}{\lambda_x} \right)^{\frac{1}{\gamma_c}}.$
David López-Pérez (M’12–SM’17) is a Senior Research Scientist and a Distinguished Member of Technical Staff (DMTS) at Nokia Bell Laboratories, and his main research interests are on small cells, ultra-dense networks and unlicensed spectrum technologies, where he has pioneered work on LTE and Wi-Fi interworking. David is currently working on massive MIMO, future indoor networks and the next generation of Wi-Fi technology. IEEE802.11be. David has authored more than 135 research articles, holds over 49 patents applications, and has received a number of prestigious awards. He is an editor of IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS.

Adrian García-Rodríguez (S’13–M’17) is a Research Scientist in Nokia Bell Labs (Ireland), where he focuses on the design of next-generation 802.11 technologies and UAV communications. He joined Bell Labs in 2016, after receiving the Ph.D. degree in Electrical and Electronic Engineering from University College London (U.K.). Previously, he held research positions in the research institute for technological development and Communication Innovation (IDetIC) at the University of Las Palmas de Gran Canaria (Spain) between 2010–2012, and in the RF group of Nokia Bell Labs (Ireland) in 2015. Adrian is co-inventor of fourteen filed patent families and co-author of 40+ technical publications in the areas of wireless communications and networking. On these topics, he frequently delivers technical tutorials (IEEE WCNC’18, IEEE ICC’18, IEEE PIMRC’18, and IEEE Globecom’18), organizes workshops and special sessions (IEEE Globecom’17, Asilomar’18, and IEEE ICC’19), and participates in industrial seminars (IEEE Globecom’18, and IEEE ICC’19). Adrian was named an Exemplary Reviewer for IEEE COMMUNICATIONS LETTERS in 2016, and both IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS and IEEE TRANSACTIONS ON COMMUNICATIONS in 2017.

Giovanni Geraedts (S’11–M’14) is an Assistant Professor and Junior Leader Fellow at Universitat Pompeu Fabra in Barcelona (Spain). He earned a Ph.D. degree from the University of New South Wales (Australia) in 2014. He gained industrial innovation experience at Nokia Bell Labs (Ireland), where he was a Research Scientist in 2016-2018. His background also features appointments at the Singapore University of Technology and Design (Singapore) in 2014-2015, the University of Texas at Austin (USA) in 2013, Supélec (France) in 2012, and Alcatel-Lucent (Italy) in 2009. He has co-authored 50+ IEEE publications with 1000+ citations, and is co-inventor of a dozen filed patent families on wireless communications and networking. He is deeply involved in the research community, serving as an editor for the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS and IEEE COMMUNICATIONS LETTERS, and as a workshop or special session co-organizer at IEEE Globecom’17, Asilomar’18, and IEEE ICC’19. He is also a frequent speaker and his contributions include a workshop keynote at IEEE PIMRC ’18, an industry seminar at IEEE ICC’19, and tutorials at IEEE WCNC’18, IEEE ICC’18, IEEE Globecom’18, and IEEE PIMRC’19. He is the recipient of the IEEE ComSoc Outstanding Young Researcher Award for Europe, Middle-East & Africa 2018.

Paolo Baracca (S’11–M’13) is a Research Engineer at Nokia Bell Labs in Stuttgart, Germany, since 2013. He received the B.Sc. and the M.Sc. degree in Telecommunications Engineering in 2007 and 2009, respectively, and the Ph.D. degree in Information Engineering in 2013, all from the University of Padova, Italy. His research interests include signal processing, multi-antenna techniques and scheduling for wireless communications. He has co-authored more than 30 research papers, holds more than 20 issued or pending patents, and regularly serves as reviewer of IEEE journals and conferences.

M. Majid Butt (S'07–M'10–SM'15) received the M.Sc. degree in digital communications from Christian Albrechts University, Kiel, Germany, in 2005, and the Ph.D. degree in telecommunications from the Norwegian University of Science and Technology, Trondheim, Norway, in 2011. He is a Senior Scientist 5G+ Research at Nokia Bell Labs, Paris-Saclay, France, and a Visiting Research Assistant Professor at Trinity College Dublin, Dublin, Ireland. Prior to that, he has held various positions at the University of Glasgow, Glasgow, U.K., Trinity College Dublin, Fraunhofer HHI, Berlin, Germany, and the University of Luxembourg, Luxembourg City, Luxembourg. His current research interests include communication techniques for wireless networks with a focus on radio resource allocation, scheduling algorithms, energy efficiency, and machine learning for RAN. He has authored more than 60 peer-reviewed conference and journal publications in these areas. Dr. Butt was a recipient of the Marie Curie Alain Bensoussan Post-Doctoral Fellowship from the European Research Consortium for Informatics and Mathematics (ERCIM). He has served as the Organizer/Chair for various technical workshops on various aspects of communication systems in conjunction with major IEEE conferences, including Wireless Communications and Networking Conference, Globecom, and Greencom. He has been an Associate Editor for the IEEE ACCESS and the IEEE COMMUNICATIONS MAGAZINE since 2016.

Nicola Marchetti (M’13–SM’15) is currently Assistant Professor in Wireless Communications at Trinity College Dublin, Ireland. He performs his research under the Irish Research Centre for Future Networks and Communications (CONNECT), where he leads the Wireless Engineering and Complexity Science (WhyCOM) lab. He received the Ph.D. in Wireless Communications from Aalborg University, Denmark in 2007, and the M.Sc. in Electronic Engineering from University of Ferrara, Italy in 2003. He also holds an M.Sc. in Mathematics which he received from Aalborg University in 2010. His collaborations include research projects in cooperation with Nokia Bell Labs and US Air Force Office of Scientific Research, among others. His research interests include Radio Resource Management, Self-Organising Networks, Complex Systems Science and Signal Processing for communication networks. He has authored in excess of 120 journals and conference papers, 2 books and 8 book chapters, holds 3 patents, and received 4 best paper awards. He is a senior member of IEEE and serves as an associate editor for the IEEE INTERNET OF THINGS JOURNAL since 2018 and the EURASIP JOURNAL ON WIRELESS COMMUNICATIONS AND NETWORKING since 2017.