

RESEARCH

Open Access



# Statistical multiplexing and pilot optimization in fronthaul-constrained massive MIMO

Jay Kant Chaudhary<sup>1\*</sup> , Jens Bartelt<sup>1,2</sup> and Gerhard Fettweis<sup>1</sup>

## Abstract

Cloud-radio access network (C-RAN) has been an attractive solution in the recent years for the future generation mobile networks due to its promising benefits. However, the transport link between the remote radio unit (RRU) and the baseband unit (BBU), known as fronthaul (FH), imposes stringent requirements in terms of data rate, latency, jitter, and synchronization. In the conventional C-RAN, the FH capacity scales linearly with the number of the transmitting antennas, which has posed severe demands on the FH capacity, especially due to emerging 5G technologies such as massive MIMO. However, this can be relaxed by performing precoding at the RRUs instead of centrally at BBU, leading to FH traffic which depends on the number of currently served users. This paper adapts queueing model and spatial traffic model to exploit randomness of the user traffic to achieve statistical multiplexing gain. Through this, we showed that the required FH capacity can be reduced significantly, depending on traffic demand and its statistical properties. Furthermore, we analyzed the impacts of pilots on capacity-constrained FH.

**Keywords:** Cloud-radio access network (C-RAN), Fronthaul, Massive MIMO, Outage probability, Pilot optimization, Queueing model, Statistical multiplexing, Traffic model

## 1 Introduction

### 1.1 Cloud radio access networks

The wireless industry has witnessed phenomenal growth over the recent years. Cisco predicts that the mobile data traffic will increase sevenfold from 2016 to 2021 and the number of devices connected to IP networks will be three times as high as the global population in 2021 [1]. The demand of higher data and continuous connectivity along with the quality of experience (QoE) is increasing daily. Moreover, plethora of use cases and application scenarios [2] envisioned for 5G such as enhanced mobile broadband (eMBB) services, machine-type communications (mMTC), and ultra-reliable and low latency communications (uRLLC) have imposed significant challenges on the radio access network (RAN) of state-of-the-art 4G technologies, thus leading to the evolution of RAN. Along with RAN's evolution, cloud computing has emerged as a popular paradigm due to its attractive characteristics such as

resource sharing, virtualization, and flexibility [3]. Hence, inheriting the merits of cloud computing into RAN, cloud RAN (C-RAN) has been proposed as a new baseband architecture for the mobile networks of the future [3].

In traditional networks, basebands units (BBUs) and analog front ends are co-located at the base station (BS) site as a single BS entity, where the BS performs the complete baseband processing including physical (PHY) layer, media access control (MAC) layer, and a part of the network layer processing. However, this setup suffers from several limitations such as higher capital expenditures (CAPEX) and operating expenses (OPEX), increased energy consumption, underutilized dedicated resources, and limited flexibility [4]. In order to address these limitations, cloud-radio access network (C-RAN) has been proposed [5, 6]—differently from conventional RAN—as a prospective architecture, where the BBU is decoupled from the RRU. That means, in C-RAN, the main digital baseband processing is centralized to the BBU, whereas a part of the signal processing functionalities of traditional BSs are moved to the remote radio units (RRUs), which

\*Correspondence: [jay\\_kant.chaudhary@tu-dresden.de](mailto:jay_kant.chaudhary@tu-dresden.de)

<sup>1</sup>Vodafone Chair Mobile Communications Systems, Technische Universität, Dresden, Germany

Full list of author information is available at the end of the article

perform the basic signal transmission and reception functionalities such as amplification and filtering.

C-RAN has drawn significant attention in the recent years and offers several advantages compared to the conventional RAN [3, 7]: Firstly, by utilizing centralized signal processing in the BBU pool instead of the distributed BSs in the conventional RAN, C-RAN significantly lowers the CAPEX and OPEX, simplifies repair and maintenance, and eases system upgrades. Secondly, joint processing in the BBU pool and cooperative radio techniques (e.g., coordinated multipoint (CoMP), enhanced inter-cell interference coordination (eICIC)) over RRUs aims at improving the spectrum efficiency, link reliability, and the communication quality, particularly of the cell edge users. Thirdly, centralized BBU pool helps to reduce the power consumption and enables efficient hardware utilization through resource sharing and network function virtualization (NFV).

In C-RAN, the BBU and RRU are generally connected by a high-speed transport link known as fronthaul (FH) that forwards the digitized I/Q (in-phase/quadrature-phase) samples from RRUs to BBUs and vice versa using common public radio interface (CPRI) protocol [8]. CPRI is a serial bidirectional digital interface, and currently, it is the most adopted specification for the FH interface implementation. Other specifications are open base station architecture initiative (OBSAI) and open radio interface (ORI) [9]. Although FH commonly employs dedicated fiber as a digital radio over fiber (D-RoF) solution, recently, radio over Ethernet (RoE) [10] is also broadly investigated [11, 12]. On the other hand, analog radio over fiber (A-RoF) solutions have not yet been standardized [11].

## 1.2 Challenges in fronthaul and related work

C-RAN not only offers many attractive advantages but also imposes several challenging requirements in FH links. In order to enable efficient centralized and cooperative processing, FH links must offer huge bandwidth, very low latency and jitter, and very tight synchronization. Unfortunately, the practical FH is often capacity-constrained or time-delay constrained [13]. Furthermore, FH networks mainly use dedicated, point-to-point fiber links with a very high capacity, which make FH networks inflexible and expensive to deploy. Thus, if the FH networks are not dimensioned correctly, they could become a bottleneck for the performance of the future mobile networks [14].

Despite promising advantages of C-RAN mentioned earlier, one of the major challenges of C-RAN is to meet its high fronthaul capacity demand. In general, capacity of a fronthaul network (in bits/s) is given by [15]:

$$D_{FH} = N_A \cdot f_s \cdot N_Q \cdot 2 \cdot \gamma, \quad (1)$$

where  $N_A = N_a \cdot N_s$  is the total number of antennas at the RRU, with  $N_a$  being the number of antennas/sector and  $N_s$

being the number of sectors,  $f_s$  is the sampling frequency (in samples/s/carrier),  $N_Q$  is the number of quantization resolution bits of ADC or DAC (in bits/sample), the factor 2 is multiplication factor for the I/Q samples, and  $\gamma = \gamma_{CW} \cdot \gamma_{LC}$  is the CPRI specific overhead factor, where  $\gamma_{CW}$  represents overhead introduced by CPRI control words and  $\gamma_{LC}$  represents line coding overhead (e.g., 10/8 for 8B/10B coding or 66/64 for 64B/66B coding).

One of the biggest challenges on C-RAN arises from the massive MIMO, which is considered as a promising technology for the future next-generation wireless networks. In massive MIMO [16–18], a BS with hundreds of antenna elements serves tens of users in the same time-frequency resources and one of the main ideas behind massive MIMO is to provide BSs with large spatial multiplexing gains and beamforming capabilities. By simultaneously serving several users, massive MIMO increases spectral efficiency. Similarly, by allowing multiple antennas to focus the radiated energy into even smaller regions of space high-energy efficiency can be achieved. Since in (1) the FH capacity scales linearly with the number of transmitting antennas  $N_A$ , with the massive MIMO system that employs arrays of hundreds of transceivers, FH data rate increases dramatically; hence, an FH network with existing dedicated optical fiber becomes very expensive to deploy. For example, for one LTE sector with 20 MHz carrier and  $2 \times 2$  MIMO,  $N_A = 2$ ,  $f_s = 30.72$  MHz,  $N_Q = 15$ ,  $\gamma_{CW} = 16/15$ , and  $\gamma_{LC} = 10/8$ , FH demands approximately 2.45 Gbps using CPRI line rate option 3 [8]. However, in massive MIMO systems with 256 antennas, as an example, this data rate would increase to about 310 Gbps. For comparison, the most recent CPRI standard specifies line rates of only up to 24 Gbps.

In order to ease the challenging requirement of FH data constraint, various solutions have been proposed such as (i) increasing the FH capacity using single fiber bidirection (SFB), wavelength division multiplexing (WDM) [19, 20], and time-shared optical networks (TSON) [21, 22]; (ii) decreasing the FH capacity by using compression techniques such as non-linear quantization and IQ data compression with a lossless 2:1 compression ratio [23], or using new functional splits between RRU-BBU [15, 24]. Another approach is packet-based network such as Ethernet which is heavily investigated by the recent initiative such as IEEE 1914 working group for next-generation fronthaul interface (NGFI) [25].

## 1.3 Contributions

In this paper, we consider a packetized network to analyze possible fronthaul capacity reduction. Recently, CPRI released the first eCPRI specification (1.0) [26] that enables use of packet-based transport technologies such as Ethernet and supports *real-time traffic* through different flexible functional splits. Various functional

splits between BBU-RRU have been investigated, e.g., in [14, 15, 24] and the choice of a particular split is an application specific. Unlike packetized network, in traditional CPRI, FH data rate is always *static* and *independent* of the traffic load, i.e., full-FH data rate needs to be forwarded even when there is no user connected to the BS. However, with the appropriate RRU-BBU functional split such as Split C<sup>1</sup> [14], where resource mapping and precoding operations are executed at RRU instead of centrally at BBU, FH data rate can be made traffic dependent. This allows FH data rate more closely coupled with the actual user traffic i.e., the traffic will be lower due to low demand or due to unfavourable channel conditions. For simplicity, we assume that the RRU generates the beamforming weights locally at the RRU after having obtained perfect CSI from uplink pilots, which means there is generally no signalling overhead on the FH. Precoding at RRU enables to transmit one stream per user instead of one stream per transceiver<sup>2</sup>. Performing precoding at RRU gives rise to two advantages: First, the number of streams will vary according to the users currently served, and hence, by allowing a certain outage probability within the limits of acceptable quality of service (QoS), i.e. dimensioning the FH capacity only for the 99th percentile of the traffic distribution, the required FH capacity can be reduced considerably. Second, the variable streams of different RRUs can be combined in the aggregation segment, resulting in statistical multiplexing, which further lowers the required FH capacity. In Section 3, we explain this concept.

Since the traffic is varying, randomness of the user traffic could be exploited to enjoy the statistical multiplexing gain. The possible factors to exploit statistical multiplexing gain are variable FH streams, aggregation of the transport streams from different cells and a reasonable outage probability. In the paper [27], we have combined these factors to obtain statistical multiplexing gain using queueing model and spatial the traffic model and are explained in Section 4. Furthermore, in this paper, we analyze the impacts of pilots on statistical multiplexing gain using the similar system model described in [27]. Since the number of pilots that can be assigned to the active users is limited, we will analyze the impacts of pilots on capacity-constrained FH. We call this optimization approach as *pilot-based optimization* and through this we will show that additional reduction in required FH capacity can be achieved as explained in Section 5.

## 2 Methods/experimental

Our aim is firstly, to study the statistical multiplexing in capacity-constrained C-RAN and secondly, to analyze impact of pilots on statistical multiplexing gain. In order to demonstrate that, a packetized fronthaul network was considered, and we adopted the spatial traffic model and  $M/G/m/m$  queueing model to achieve the

mentioned goals. We used traffic models to get random spatial traffic maps via log normally distributed random fields defined by statistical parameters, and queueing model to get the steady state probabilities of users served by each cell. The simulations were performed using MATLAB and the simulation parameters are listed in Table 1. We assumed that with an appropriate functional split, FH data rate is coupled with the actual user traffic, hence by assuming a reasonable outage probability within an acceptable QoS, we can achieve statistical multiplexing. Further, since the number of pilots that can be assigned to users is limited, we present an optimization algorithm to adapt the number of pilots to the FH capacity. The system model considered in this work is described in detail in Section 4 and the simulation results are presented in Section 6.

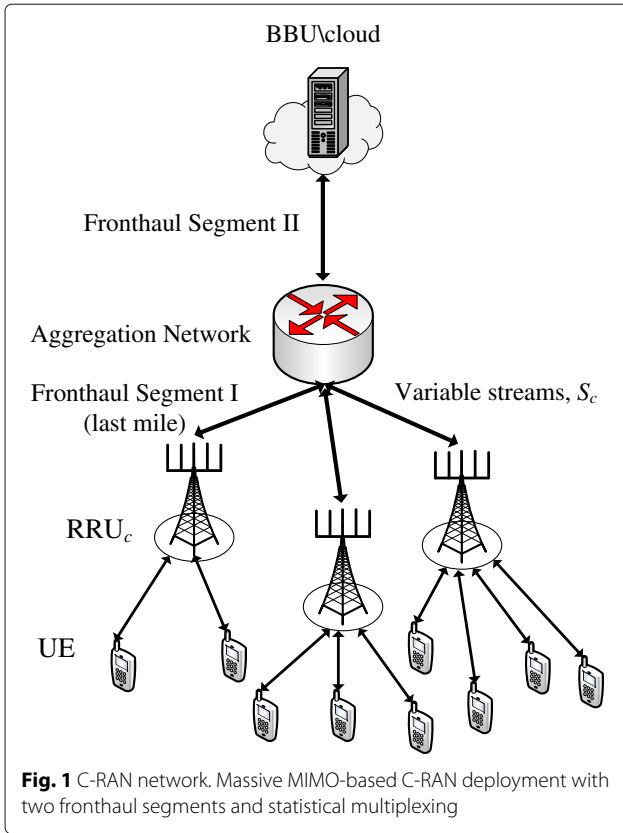
## 3 Statistical multiplexing

Figure 1 illustrates a basic network setup that enables statistical multiplexing considered in this paper. In this figure, traffics from multiple RRUs are aggregated to an *aggregation network* and the resulting traffic is then forwarded to the BBU via FH Segment II in the uplink and vice versa in the downlink [28, 29], ([30], Chapter 4). The direct link between the RRU and aggregation network is termed as *last mile*, denoted by *FH Segment I*, whereas the link between the aggregation network and the BBU is the main FH segment, denoted by *FH Segment II*. The advantage of having FH Segment I is that it allows users to have data delivery with shorter cable lengths, which otherwise would have been a single dedicated fiber between each RRU and BBU. On the other hand, FH Segment II requires higher capacity and more protection against the link failure.

Since the streams from users are varying, we can observe two methods to lower the required FH capacity:

**Table 1** Simulation parameters

Parameters	Symbol	Value
Number of cells	$C$	19
Inter site distance	$d_{ISD}$	200 m
RRH height	$h_{RRU}$	12 m
Bandwidth	$B$	20 MHz
Coherence bandwidth	$B_{coh}$	200 kHz
Coherence time	$T_{coh}$	5 ms
Number of transmitting antennas	$M_c$	256
Maximum users	$K_{max}$	64
Total transmit power	$pM_c$	23 dBm
Total noise power	$\sigma^2$	-96 dBm
Average file size	$s$	80 Mbit
Pilot reuse factor	$\beta$	1,4



first, accepting a certain outage (which benefits both FH Segments I and II), and second, accounting for the effect of statistical multiplexing in the aggregation part (which is only possible in Segment II). In this paper, we will describe a methodology how to evaluate these gains and quantify the benefits for different scenarios.

## 4 System model

### 4.1 MIMO rate

Figure 2 shows the frame structure of a massive MIMO system [31]. Assuming the channel is frequency flat with coherence time,  $\tau_{\text{coh}}$  and coherence bandwidth  $B_{\text{coh}}$ , length of the channel coherence interval (in symbols) is  $T_c = \tau_{\text{coh}} B_{\text{coh}}$ . Out of these  $T_c$  symbols,  $\tau$  symbols are dedicated for pilot signaling, and the remaining  $(T_c - \tau)$  symbols used for payload data, where  $\zeta^{(ul)}$  and  $\zeta^{(dl)}$  are fractions of UL and DL data transmissions, respectively. Massive MIMO systems commonly operate in TDD mode, where it is assumed that channel is reciprocal meaning the BS estimates the downlink channel using the uplink pilots. The number of orthogonal pilots that can be assigned to users is always limited. If FDD is used, channel state information (CSI) corresponding to UL and DL will be different since UL and DL use different frequency bands. In our model, for simplicity, we have assumed TDD operation. We also focus on the DL and assume

for simplicity that  $\zeta^{(dl)} = 1$ , i.e., there is only DL data transmission and in the UL only pilots are transmitted.

Let us now consider downlink transmission of a massive MIMO system, where the cellular network consists of a set  $\mathcal{C} = \{1, 2, \dots, C\}$  of  $C$  RRUs. Each cell has its own RRU that can simultaneously serve  $K_{\text{max}}$  users. We suppose each RRU is equipped with  $M_c$  antennas, while each user has a single antenna. Usually for massive MIMO system, it is assumed that the number of antennas at each RRU is much larger than the number of served users, i.e.,  $M_c \gg K_c \gg 1$ . We assume all the BSs and UEs are perfectly synchronized, the UEs share the same time-frequency resource, and transmissions take place over the Rayleigh fading channels.

The area served by the RRUs is denoted  $\mathcal{A}$ , with a single location indicated by its coordinates  $(x, y)$ . The pathloss factor, defined here as a ratio of received power to transmitted power, between RRU  $c$  and location  $(x, y)$  is denoted  $\alpha_c(x, y)$  and modeled according to the urban microcellular pathloss model defined in [32]. Users are associated with the RRU providing the lowest pathloss; hence, the serving area  $\mathcal{A}_c$  of a RRU  $c$  is given as

$$(x, y) \in \mathcal{A}_c \text{ if } c = \arg \max_c \alpha_c(x, y). \quad (2)$$

Let us consider that the total transmit power of a BS is  $pM_c$ , where  $p$  is the average power per antenna that is considered to be same for all the antennas. Each cell  $c$  receives interference from the active antennas in the any other cell  $d$ . Let  $M_d$  be the active antennas in any other cell; hence, total transmitting power of the corresponding cell is  $pM_d$ . Then, the signal-to-noise-plus-interference ratio (SINR) at a location  $(x, y)$  can be obtained as:

$$\gamma(x, y) = \frac{pM_c \alpha_c(x, y)}{\sigma^2 + \sum_{d \in \mathcal{C} \setminus c} pM_d \alpha_d(x, y)}, \quad (3)$$

where  $\sigma^2$  denotes the noise power.

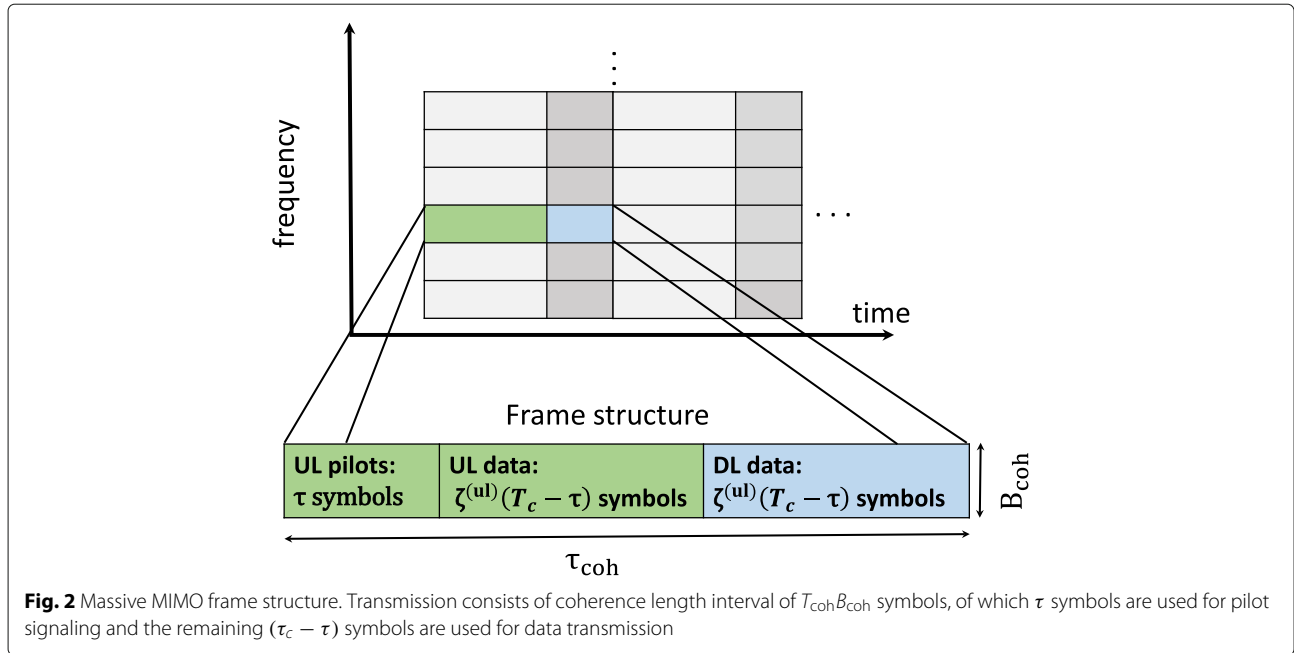
We use Poisson point process to describe the traffic demand in the network, compare, e.g. [33]. For this, we define for each location a user arrival rate per area  $\Lambda(x, y)$  (in  $1/\text{s}/\text{km}^2$ ), and a corresponding traffic density (in  $\text{Mbps}/\text{km}^2$ ) is

$$\Omega(x, y) = \Lambda(x, y) \cdot s, \quad (4)$$

where  $s$  (in bits) is the mean file size requested per user. The mean traffic density of the overall area  $\mathcal{A}$  we denote  $\bar{\Omega}$ . For the serving area of a RRU, this results in user arrivals with arrival rate (in  $1/\text{s}$ ),

$$\lambda_c = \int_{\mathcal{A}_c} \Lambda(x, y) dx dy. \quad (5)$$

From this, we define the average SINR in the serving area of RRU  $c$  as the expected value of the SINRs weighted according to the traffic distribution  $\Omega(x, y)$ , i.e.:



$$\bar{\gamma}_c = \mathbb{E}[\gamma(x, y)] = \frac{\int_{\mathcal{A}_c} \gamma_c(x, y) \Omega(x, y) dx dy}{\int_{\mathcal{A}_c} \Omega(x, y) dx dy}. \quad (6)$$

Furthermore, for simplicity, let us assume that each RRU has obtained perfect CSI from its users and employs zero forcing precoding in order to cancel out the intracell interference and adapts power allocation such that each of the  $K_c$  user achieves the same average data rate,  $R_c$  (in bps) given by [34]:

$$R_c(K_c) = B \left(1 - \frac{\beta K_{\max}}{T_c}\right) \log_2 \left(1 + \frac{\bar{\gamma}_c}{K_c} (M_c - K_c)\right), \quad (7)$$

where  $B$  is channel bandwidth,  $\beta$  is reuse factor,  $T_c = \tau_{\text{coh}}B_{\text{coh}}$  is length of channel coherence interval, and  $K_{\max}$  is the maximum number of users, which is, for now, assumed to be same for all cells.  $M_c - K_c$  is the effective array gain, and the factor  $1/K_c$  accounts for the fact that the total transmit power is split between all users. The pre-log factor  $\left(1 - \frac{\beta K_{\max}}{T_c}\right)$  corresponds to the pilot overhead, and it will play an important role, which is explained in Section 5.

Exemplary plots of the average data rate per user  $R_c$  and total sum throughput  $K_c \cdot R_c(K_c)$  are shown in Fig. 3a, b, respectively. It is obvious from Fig. 3a that higher per user average rate is achievable with fewer active users. On the other hand, Fig. 3b shows that the sum throughput increases when increasing the number of active users which highlights the general benefit of massive MIMO in terms of capacity. The simulation parameters for all the figures are listed in Table 1.

#### 4.2 Queueing model

In order to achieve statistical multiplexing gain, we need to find probabilities of random number of users served by each cell. For this, we utilize queueing theory results from [35, 36], where each MIMO RRU is modeled as  $M/G/m/m$  state-dependent queue. The  $M/G/m/m$  queue states that for exponential arrival and general distribution of service time, maximum  $m$  number of users can be served simultaneously ( $m$  servers and no waiting) and the arrivals follow the Poisson process [34]. In this work, we assume that the maximum number of users that a BS can serve is  $m = K_{\max}$ . Let  $\pi_c(n) \equiv P_r[K_c = n]$  be the steady state probabilities of having  $K_c$  active users served by RRU  $c$ , then under  $M/G/m/m$  state-dependent queue,  $\pi_c(n)$  is given by [35]:

$$\pi_c(n) = \left[ \frac{\left[\lambda_c \frac{s}{R_c(1)}\right]^n}{n! f(n) f(n-1) \dots f(2) f(1)} \right] \pi_c(0), \quad (8)$$

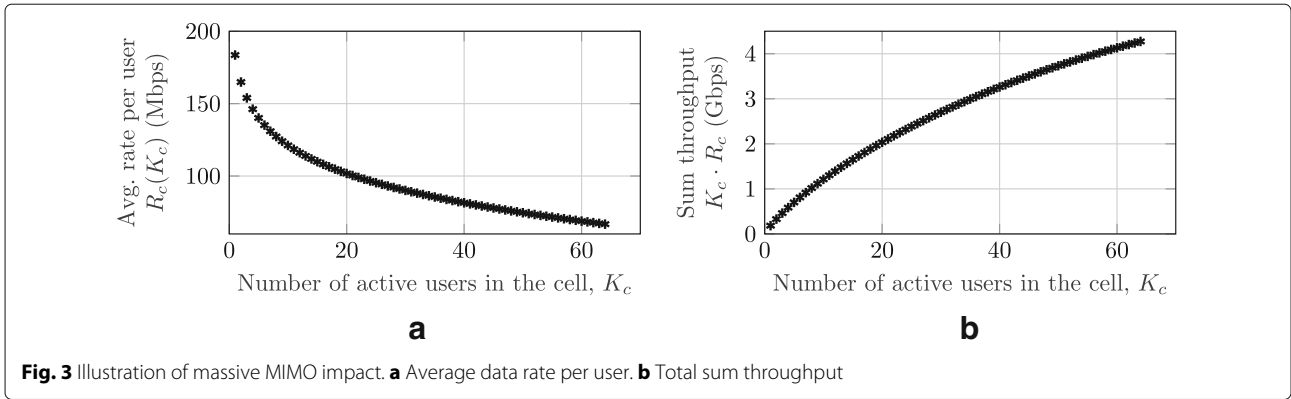
$$n = 1, 2, \dots, m,$$

with

$$\pi_c^{-1}(0) = 1 + \sum_{i=1}^m \left( \frac{\left[\lambda_c \frac{s}{R_c(1)}\right]^i}{i! f(i) f(i-1) \dots f(2) f(1)} \right),$$

where  $f(n) = R_c(n)/R_c(1)$  is the normalized rate per user,  $R_c$  is the average data rate per user while serving  $n$  number of users given by (7),  $\lambda_c$  is the arrival rate from (5), and  $s$



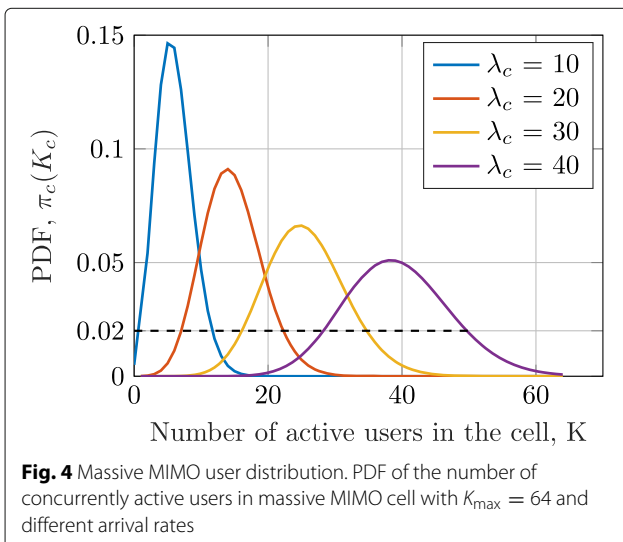


is the average file size requested by each user introduced earlier.

An example of user distributions defined by (8) is given in Fig. 4 for different values of  $\lambda_c$ . Figure 4 implies that as the arrival rate increases, there are more flows per second from the users and the number of users attempting to get the resources is also increasing. For example, at 2% probability, the number of active users for arrival rate of  $\lambda_c = 10$  flows per second is 12 which increases to 50 users for arrival rate of  $\lambda_c = 40$  flows per second.

### 4.3 Fronthaul capacity, outage, and multiplexing

In literature, there are two main different definitions of the FH capacity [3]: The first one states that FH capacity is defined as the maximum sum data rate transmitted on each FH. In this case, the authors always implicitly assume each FH can serve unlimited number of users. However, this assumption can not hold in real systems. On the other hand, the second definition states that FH capacity is defined as maximum number of users that can be served on each FH. In this paper, we adopt the second definition.



Conventionally, the FH in Segment I would be dimensioned to serve its maximum number of users, i.e. for a number of stream  $S_{c,\max} = K_{\max}$  streams. Similarly, Segment II would be dimensioned for  $S_C = C \cdot K_{\max}$  streams. Such a dimensioning is common in conventional CPRI-based FH networks, which requires a static and constant data rate per RRU. This high and constant data rate is the major challenge in massive MIMO C-RANs as explained in Section 1.2. From network operators perspective, it would be beneficial for them to constrain that capacity to lower deployment cost. As the traffic is varying due to the varying number of user streams, we can assume a certain outage probability  $P_O$  on each link according to some QoS requirements. Hence, the FH in Segment I can be dimensioned with the outage capacity

$$S_{c,O} = n \text{ such that } \Pi_c(n) < P_O, \tag{9}$$

where  $\Pi_c(n) = \sum_{i \leq n} \pi_c(i)$  is CDF of each individual cell. Furthermore, the streams are aggregated for Segment II, i.e., the number of streams are summed up. The summation leads to a convolution of the corresponding probability distributions, i.e., the distribution of user streams on Segment II becomes:

$$\pi_C = \pi_1 * \pi_2 * \dots * \pi_c, \tag{10}$$

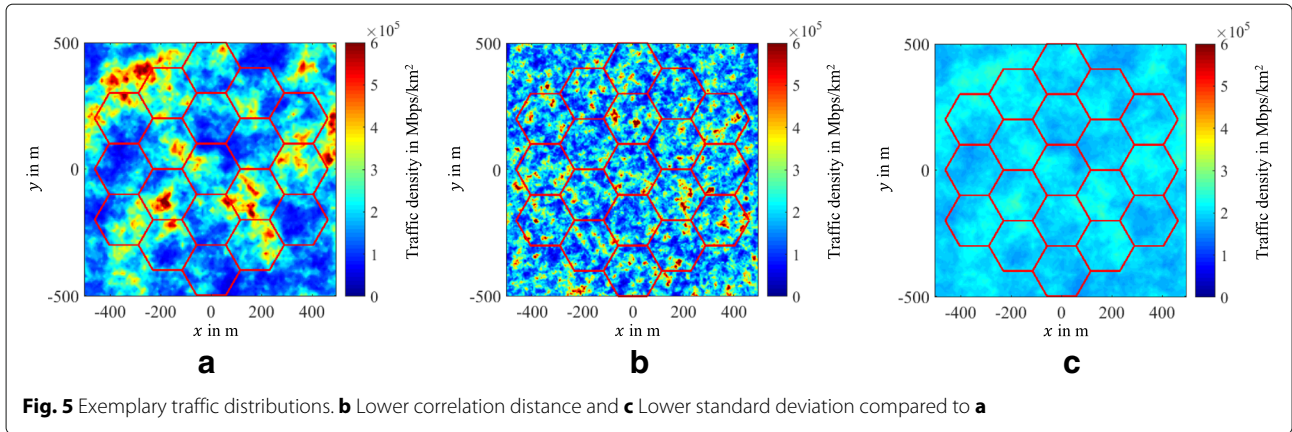
with CDF  $\Pi_C$  and outage capacity

$$S_{C,O} = n \text{ such that } \Pi_C(n) < P_O. \tag{11}$$

In general, the convolution will lead to a longer-tailed PDF, which yields to a statistical multiplexing gain, as  $S_{C,O} \leq \sum_C S_{c,O}$ . To assess the benefit of the statistical multiplexing, we define the relative required FH rates in Segment I and II as:

$$S_1 = \frac{\sum_C S_{c,O}}{C \cdot S_{c,\max}} \quad \text{for FH Segment I,} \tag{12}$$

$$S_2 = \frac{S_{C,O}}{C \cdot S_{c,\max}} \quad \text{for FH Segment II.} \tag{13}$$



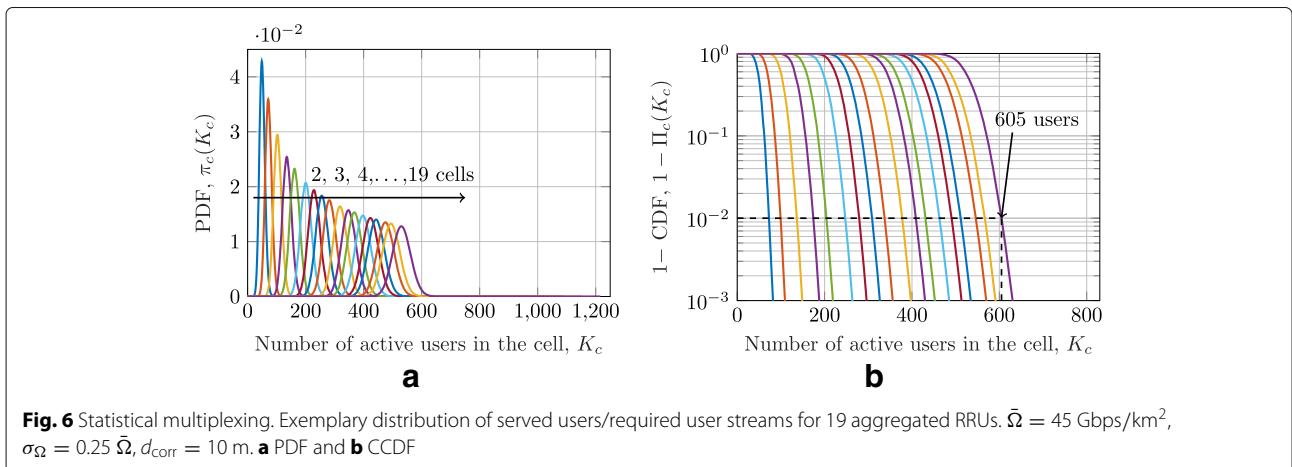
**4.4 Traffic model**

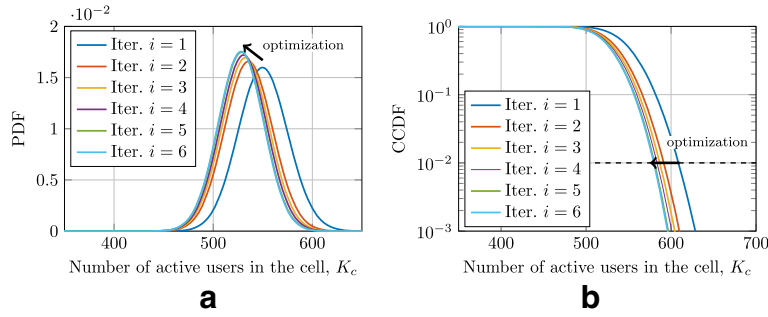
In general, the gain of statistical multiplexing will depend on the variance of the total number of streams. This variance is affected both by the (temporal) variation for users from (8) and by the different (spatial) variation of users among different cells based on the traffic distribution  $\Omega(x, y)$ . In order to model  $\Omega(x, y)$ , we utilize a traffic model developed in [37, 38]. This traffic model allows to create random spatial traffic maps via log-normal distributed random fields defined by three statistical parameters: mean traffic density ( $\bar{\Omega}$ ), traffic density standard deviation ( $\sigma_{\Omega}$ ), and a correlation distance ( $d_{corr}$ ). Three different examples of such traffic maps are given in Fig. 5. The parameter  $\bar{\Omega}$  controls the overall traffic demand,  $\sigma_{\Omega}$  controls the ratio between traffic demand in hot spots and low-traffic areas, and  $d_{corr}$  controls the size of the hotspots. With traffic maps generated based on this model, we can average statistical multiplexing gains over random scenarios without having to rely on just a single scenario, leading to more consistent results and more general conclusions for real scenarios. For more details on this traffic model see [38].

Now, in order to illustrate the underlying concepts, a layout of 19 homogeneous hexagonal cells each having inter site distance of 200 m are plotted as shown in Fig. 5. It is common practice to assume regular cells, in particular hexagonal cells to, establish the general properties, although the practical deployments have irregular cells. The innermost cell is surrounded by a tier of six cells, which in turn are surrounded by additional tier of 12 cells. Figure 6a, b respectively illustrates the PDF and CCDF of each cell. The total traffic from 19 such cells, assuming each cell in its peak load can serve 64 users, demands to have total  $64 \times 19 = 1216$  user streams to be forwarded. However, assuming a reasonable 1% outage probability of the fronthaul Segment II, we need to transport only 605 users as shown in Fig. 6b, which means less fronthaul capacity demand. This shows up to 50% FH capacity saving can be achieved.

**5 Pilot optimization**

In (7), the pre-log scaling factor  $\left(1 - \frac{\beta K_{max}}{T_c}\right)$  is the channel estimation overhead.  $K_{max}$  is the maximum number of users that a BS can support due to the number





**Fig. 7** Pilot-based optimization. Example of the pilot-based optimization for  $K_{\max} = 64$  users. **a** PDF and **b** CCDF

of transmitted pilots. This number is in general a system design parameter and would be set according to an *expected* general peak demand. However, as we have shown in the previous section in the case of FH, it can be much more efficient to design a system based on the *actual* demand. According to (7), it would increase the rate of all active users, if a lower number of pilots is used. What is even more important is that we already dimension the fronthaul to support only a limited number of users based on an acceptable outage probability. It hence does not make sense to support more users in terms of pilots if they cannot be served by the FH anyways. We can hence derive a simple, iterative optimization algorithm to adapt the number of pilots to the fronthaul capacity, which is in turn based on the outage probability. For this, we assume from now on that the number of pilots/number of supported users can be different for each RRU  $c$  and is denoted  $K_{\max,c}$ . The algorithm to find the optimal number of pilots  $K_{\max,c}^*$  is depicted in Algorithm 1.

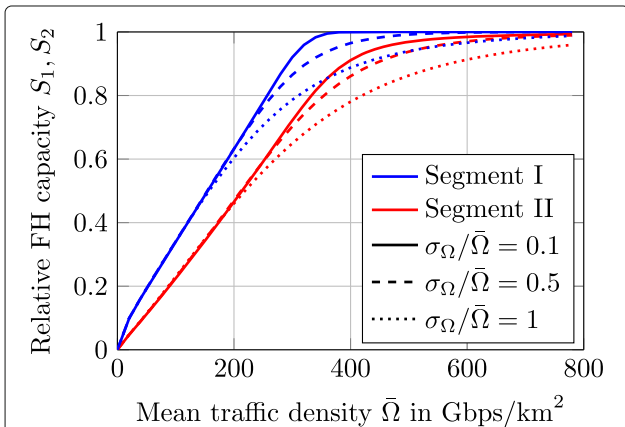
The algorithm can be explained as follows. We set the number of pilots to an starting value  $K_{\max}$  for all RRUs  $c$ . We then calculate the number of Segment I outage FH

**Algorithm 1** Pilot optimization

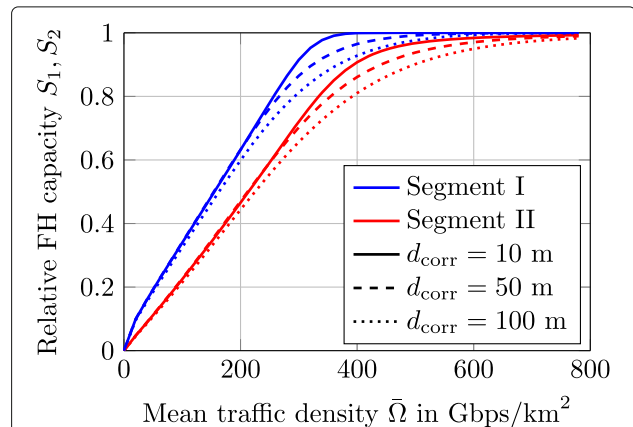
```

i = 0
K_max,c^(0) = K_max ∀ c ∈ C
repeat
  for all c ∈ C do
    calculate (7), (8), (9),(11)
    K_max,c^(i) = S_c,O
  end for
  i = i + 1
until K_max,c^(i) = K_max,c^(i-1) ∀ c ∈ C
K_max,c* = K_max,c^(i), S_c,O* = S_c,O^(i), S_c,O* = S_c,O^(i) ∀ c ∈ C
    
```

streams  $S_{c,O}$  according to (9). This is the number of FH streams/users we support for each RRU in Segment I. Now we can use that value as the new number of pilots in each RRU. This will increase the rate of all users according to (7), which in turn leads to less number of users in the queue according to (8) (illustrated in Fig. 7a) which in turn again may reduce the number of outage streams in (9) (illustrated in Fig. 7b). The algorithm terminates when the

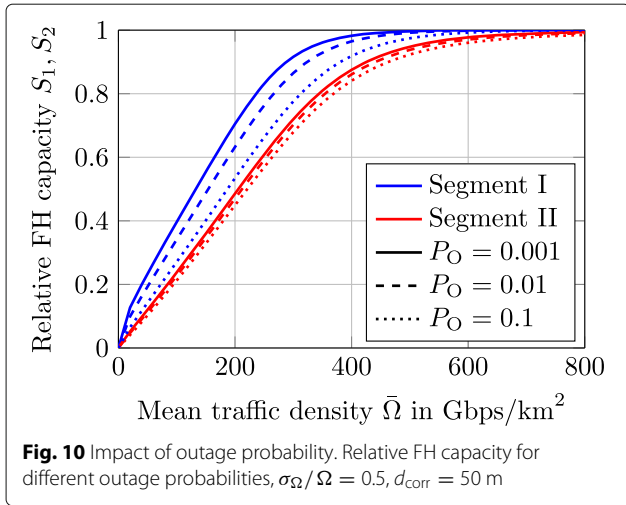


**Fig. 8** Impact of standard deviation. Relative FH capacity for different relative standard deviations of the traffic density,  $d_{\text{corr}} = 50$  m,  $P_O = 0.01$



**Fig. 9** Impact of traffic correlation. Relative FH capacity for different traffic correlation distances,  $\sigma_{\Omega}/\bar{\Omega} = 0.5$ ,  $P_O = 0.01$





increase in rate is so low that it does not lead to a reduction of  $n$  in (7) by at least one. The convergence is hence guaranteed by the limitation of  $n$  to integer values.

After the final iteration, the optimal number of pilots will be equal to the fronthaul capacity in Segment I. Of course, this also means that the outage no longer occurs in Segment I but already during user admission in the wireless link, as no more pilots are available. To illustrate the additional benefit achieved by the optimization we define the optimization gain in terms of FH capacity as:

$$g_1 = 1 - \frac{\sum_C S_{c,O}^*}{\sum_C S_{c,O}} \quad \text{for FH Segment I,} \quad (14)$$

$$g_2 = 1 - \frac{S_{c,O}^*}{S_{c,O}} \quad \text{for FH Segment II.} \quad (15)$$

## 6 Numerical results

### 6.1 Scenario

To evaluate the FH capacity reduction, we utilize an exemplary setup in Fig. 5 consisting of 19 uniformly placed hexagonal cells with inter site distance  $d_{\text{ISD}} = 200 \text{ m}$  and RRU's placed at a height of  $h_{\text{RRU}} = 12 \text{ m}$ . These cells are placed on random traffic maps generated according to

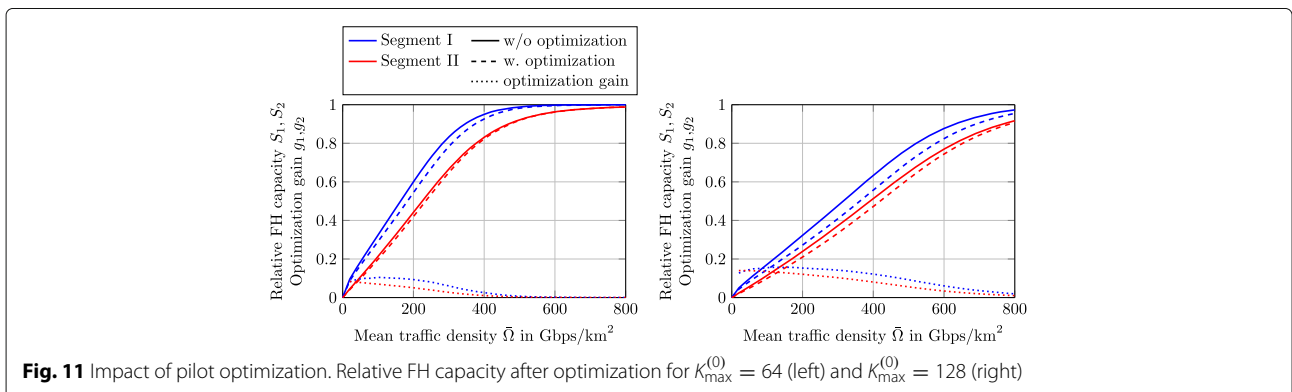
Section 4.4, which is also illustrated in Figs. 8, 9, and 10. Then, the relative required FH rates according to (12) and (13) are evaluated and the results are averaged over 25 instances of random traffic maps.

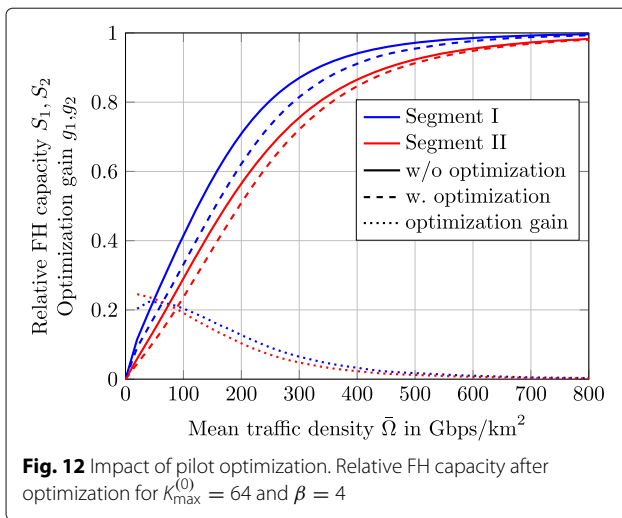
### 6.2 Statistical multiplexing

Figures 8, 9, and 10 illustrate the reduction in relative FH capacity that is achieved by accepting outage and utilizing statistical multiplexing in Segment II, first without pilot optimization. As can be seen, the relative FH capacity mainly scales with the mean traffic density. In addition, the capacity in FH Segment II is always lower, as here the additional effect of statistical multiplexing comes into effect. The difference between Segments I and II is more pronounced towards higher traffic densities. Here, clearly the statistical multiplexing effect is more dominant compared to the reduction possible by accepting outage. Furthermore, it can be seen that higher values of traffic variance and correlation distance lead to lower FH capacities, as both parameters lead to a higher variability in total cell traffic among the different RUs, hence resulting in a higher multiplexing gain. Finally, it can be seen in Fig. 10 that a higher outage probability leads to a lower required FH capacity, as can be expected. Here, especially Segment I profits. In Segment II, due to the statistical multiplexing effect, the probability distribution converges towards the mean traffic, and hence, the difference between the percentiles is less pronounced.

### 6.3 Pilot optimization

Next, Fig. 11 shows the relative FH capacity before and after pilot optimization according to Section 5. In addition, the additional reduction provided by the optimization according to (14) and (15) is illustrated. As can be seen, the optimization achieves an additional reduction in required FH capacity of up approximately 15% in both segments. As can be also seen, the gain is of course higher when the starting point  $K_{\text{max}}^{(0)}$  is chosen larger, as in this case there is more room for improvement. Furthermore, we consider pilot reuse factor  $\beta = 4$  in Fig. 12, unlike





all other simulated figures, where we assumed  $\beta = 1$  for simplicity. Higher value of  $\beta$  allows to have sufficient pilot resources to be shared among the users and hence, it helps in mitigating the pilot contamination [31]. It is to be noted that proper choice of  $\beta$  depends on various factors such as number of transmitting antennas, SINR values, number of allocated users, and spectral efficiency. As seen from Fig. 12, the optimization gain has now improved to roughly 20%.

## 7 Conclusions

The introduction of massive MIMO systems will pose severe challenges on the C-RAN's FH network, causing the FH network to be a constraint for future deployments. However, this can be mitigated by implementing per-user fronthauling and enabling statistical multiplexing in aggregation networks. By adopting queuing model and spatial traffic model, we have shown that assuming a reasonable outage probability within the limits of acceptable quality of service, user-based FH can reduce the required FH capacity significantly. We analyzed the impacts of traffic density, correlation distance, and outage probability and showed that the relative FH capacity in the FH Segment II is always lower than that in the FH Segment I. Furthermore, we have also analyzed impacts of pilots on capacity-constrained FH. We introduced an optimization scheme and showed that additional reduction in FH segments can be achieved.

## Endnotes

<sup>1</sup>The nomenclature of functional split is not harmonized and is used differently by different works.

<sup>2</sup>Investigation of impact of how and where precoding scheme has been performed is beyond the scope of the focus of this paper.

## Abbreviations

A-RoF: Analog radio over fiber; BBU: Base band unit; BS: Base station; CAPEX: Capital expenditures; CCDF: Complementary cumulative distribution function; CDF: Cumulative distribution function; CoMP: Coordinated multipoint; CPRI: Common public radio interface; C-RAN: Cloud-radio access network; CSI: Channel state information; D-RoF: Digital radio over fiber; eCPRI: enhanced CPRI; eMBB: enhanced mobile broadband; FDD: Frequency division duplexing; FH: Fronthaul; I/Q: In-phase/quadrature-phase; IEEE: Institute of electrical and electronics engineers; LTE: Long-term evolution; MAC: Medium access control; MIMO: Multiple input multiple output; mMTC: massive machine-type communication; NFV: Network function virtualization; NGFI: Next-generation fronthaul interface; OBSAI: Open base station architecture initiative; OPEX: Operating expenditures; ORI: Open radio interface; PDF: Probability density function; PHY: Physical layer; QAM: Quadrature amplitude modulation; QoS: Quality of experience; QoS: Quality of service; RAN: Radio access network; RoE: Radio over Ethernet; RoF: Radio over fiber; RRU: Remote radio unit; SINR: Signal-to-noise-plus-interference ratio; TDD: Time division duplexing; TSON: Time-shared optical networks; UE: User equipment; uRLLC: Ultra-reliable and low-latency communications; WDM: Wavelength division multiplexing

## Acknowledgements

The research leading to this work has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 671551 (5G-XHaul). The European Union and its agencies are not liable or otherwise responsible for the contents of this document; its content reflects the views of its authors only.

## Authors' contributions

All authors read and approved the final manuscript.

## Competing interests

The authors declare that they have no competing interests.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Author details

<sup>1</sup>Vodafone Chair Mobile Communications Systems, Technische Universität, Dresden, Germany. <sup>2</sup>Airrays GmbH, Dresden, Germany.

Received: 16 October 2017 Accepted: 30 July 2018

Published online: 16 August 2018

## References

1. Cisco Systems, Cisco Visual Networking Index: Forecast and Methodology, 2016–2021 (2017). Available: <https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/complete-white-paper-c11-481360.html>. Accessed 25 Sept 2017
2. NGMN Alliance, NGMN White Paper (2015). Available: [https://www.ngmn.org/fileadmin/ngmn/content/downloads/Technical/2015/NGMN\\_5G\\_White\\_Paper\\_V1\\_0.pdf](https://www.ngmn.org/fileadmin/ngmn/content/downloads/Technical/2015/NGMN_5G_White_Paper_V1_0.pdf). Accessed 2 Dec 2016
3. J Tang, WP Tay, TQS Quek, B Liang, System cost minimization in Cloud RAN with limited fronthaul capacity. *IEEE Trans. Wirel. Commun.* **16**(5), 3371–3384 (2017)
4. 5G PPP Architecture Group, View on 5G Architecture (2016). Available: [https://5g-ppp.eu/wp-content/uploads/2017/07/5G-PPP-5G-Architecture-White-Paper-2-Summer-2017\\_For-Public-Consultation.pdf](https://5g-ppp.eu/wp-content/uploads/2017/07/5G-PPP-5G-Architecture-White-Paper-2-Summer-2017_For-Public-Consultation.pdf). Accessed 25 Sept 2017
5. China Mobile Research Institute, C-RAN - The road towards green RAN. White Paper (2011). Available: [http://labs.chinamobile.com/cran/wp-content/uploads/CRAN\\_white\\_paper\\_v2\\_5\\_EN.pdf](http://labs.chinamobile.com/cran/wp-content/uploads/CRAN_white_paper_v2_5_EN.pdf). Accessed 2 Sept 2016
6. A Checko, HL Christiansen, Y Yan, L Scolari, G Kardaras, MS Berger, L Dittmann, Cloud RAN for mobile networks—a technology overview. *Commun. Surv. Tutorials IEEE.* **17**(1), 405–426 (2015)
7. T Pfeiffer, in *2015 Optical Fiber Communications Conference and Exhibition (OFC)*. Next generation mobile fronthaul architectures (Optical Society of America, 2015), pp. 1–3. <http://www.osapublishing.org/abstract.cfm?URI=OFC-2015-M2J.7>
8. CPRI, Common Public Radio Interface (CPRI); Interface Specification (V7.0). Technical report (2015). Available: [http://www.cpri.info/downloads/CPRI\\_v\\_7\\_0\\_2015-10-09.pdf](http://www.cpri.info/downloads/CPRI_v_7_0_2015-10-09.pdf). Accessed 2 Nov 2016

9. A Pizzinat, P Chanclou, F Saliou, T Diallo, Things you should know about fronthaul. *J. Light. Technol.* **33**(5), 1077–1083 (2015)
10. IEEE 1914.3 Task Force, Standard for Radio Over Ethernet Encapsulations and Mappings. Available: <http://sites.ieee.org/sagroups-1914/p1914-3/>. Accessed 28 Sept 2017
11. M Fiorani, B Skubic, J Mårtensson, L Valcarengi, P Castoldi, L Wosinska, P Monti, On the design of 5G transport networks. *Photon Netw. Commun.* **30**(3), 403–415 (2015)
12. A Haddad, M Gagnaire, in *2014 International Conference on Optical Network Design and Modeling*. Radio-over-Fiber (RoF) for mobile backhauling: A technical and economic comparison between analog and digitized RoF, (2014), pp. 132–137
13. M Peng, C Wang, V Lau, HV Poor, Fronthaul-constrained cloud radio access networks: insights and challenges. *IEEE Wirel. Commun.* **22**(2), 152–160 (2015)
14. J Bartelt, D Wübben, P Rost, J Lessmann, G Fettweis, *Fronthaul for a Flexible Centralization in Cloud Radio Access Networks*. (Wiley, New York, 2016), pp. 55–84
15. U Dötsch, M Doll, HP Mayer, F Schaich, J Segel, P Sehier, Quantitative analysis of split base station processing and determination of advantageous architectures for LTE. *Bell Labs Tech. J.* **18**(1), 105–128 (2013)
16. TL Marzetta, Noncooperative cellular wireless with unlimited numbers of base station antennas. *IEEE Trans. Wirel. Commun.* **9**(11), 3590–3600 (2010)
17. F Rusek, D Persson, BK Lau, EG Larsson, TL Marzetta, O Edfors, F Tufvesson, Scaling up MIMO: Opportunities and challenges with very large arrays. *IEEE Signal Proc. Mag.* **30**(1), 40–60 (2013)
18. EG Larsson, O Edfors, F Tufvesson, TL Marzetta, Massive MIMO for next generation wireless systems. *IEEE Commun. Mag.* **52**(2), 186–195 (2014)
19. I CL, J Huang, R Duan, C Cui, J Jiang, L Li, Recent progress on C-RAN centralization and cloudification. *IEEE Access.* **2**, 1030–1039 (2014)
20. 5G-XHaul Project, Deliverable D2.1, Requirements Specification and KPIs Document. Available: [http://www.5g-xhaul-project.eu/download/5G-XHaul\\_D\\_21.pdf](http://www.5g-xhaul-project.eu/download/5G-XHaul_D_21.pdf). Accessed 10 Feb 2017
21. GS Zervas, J Triay, N Amaya, Y Qin, C Cervelló-Pastor, D Simeonidou, in *2011 37th European Conference and Exhibition on Optical Communication*. Time shared optical network (TSON): A novel metro architecture for flexible multi-granular services (OSA, 2011), pp. 1–3. <http://www.opticsexpress.org/abstract.cfm?URI=oe-19-26-B509>
22. Y Yan, Y Qin, G Zervas, B Rofoe, D Simeonidou, in *2012 38th European Conference and Exhibition on Optical Communications*. High performance and flexible FPGA-based time shared optical network (TSON) metro node (OSA, 2012), pp. 1–3. <http://www.opticsexpress.org/abstract.cfm?URI=oe-21-5-5499>
23. J Lorca, L Cucala, in *2013 IEEE 14th International Symposium on "A World of Wireless, Mobile and Multimedia Networks" (WoWMoM)*. Lossless compression technique for the fronthaul of LTE/LTE-advanced cloud-RAN architectures, (2013), pp. 1–9
24. D Wübben, P Rost, JS Bartelt, M Lalam, V Savin, M Gorgoglione, A Dekorsy, G Fettweis, Benefits and impact of cloud computing on 5G signal processing: flexible centralization through cloud-RAN. *IEEE Signal Proc. Mag.* **31**(6), 35–44 (2014)
25. Next Generation Fronthaul Interface (1914) Working Group. <http://sites.ieee.org/sagroups-1914/>. Accessed 28 Sept 2017
26. eCPRI, Common Public Radio Interface (CPRI); eCPRI Interface Specification (V1.0). Technical report (2017). Available: <http://www.cpri.info/spec.html>. Accessed 6 Oct 2017
27. JK Chaudhary, J Bartelt, G Fettweis, in *2017 European Conference on Networks and Communications (EuCNC)*. Statistical multiplexing in fronthaul-constrained massive MIMO, (2017), pp. 1–6
28. J Liu, S Xu, S Zhou, Z Niu, Redesigning fronthaul for next-generation networks: beyond baseband samples and point-to-point links. *IEEE Wirel. Commun.* **22**(5), 90–97 (2015)
29. JP Rost, D Wübben, J Lessmann, B Melis, G Fettweis, Fronthaul and backhaul requirements of flexibly centralized radio access networks. *IEEE Wirel. Commun.* **22**(5), 105–111 (2015)
30. KMS Huq, J Rodriguez, *Backhauling/Fronthauling for Future Wireless Systems*. (Wiley, New York, 2016)
31. E Björnson, EG Larsson, M Debbah, Massive MIMO for maximal spectral efficiency: How many users and pilots should be allocated? *IEEE Trans. Wirel. Commun.* **15**(2), 1293–1308 (2016)
32. ITU-R, ITU Recommendation ITU-R M.2135-1: Guidelines for Evaluation of Radio Interface Technologies for IMT-Advanced. ITU-R. Available: [https://www.itu.int/dms\\_pub/itu-r/otp/rep/R-REP-M.2135-1-2009-PDF-E.pdf](https://www.itu.int/dms_pub/itu-r/otp/rep/R-REP-M.2135-1-2009-PDF-E.pdf). Accessed 27 Sept 2017
33. F Baccelli, B Blaszczyzyn, *Stochastic Geometry and Wireless Networks, Volume I - Theory*. Foundations and Trends in Networking Vol. 3: No 3–4, pp 249–449, vol. 1. (NoW Publishers, Delft 2009), p. 150. *Stochastic Geometry and Wireless Networks, Volume II - Applications*; see <http://hal.inria.fr/inria-00403040>. <https://hal.inria.fr/inria-00403039>
34. MMA Hossain, C Cavdar, E Björnson, R Jäntti, in *2015 IEEE Globecom Workshops (GC Wkshps)*. Energy-efficient load-adaptive massive MIMO, (2015), pp. 1–6
35. FRB Cruz, JM Smith, Approximate analysis of M/G/c/c state-dependent queueing networks. *Comput. Oper. Res.* **34**(8), 2332–2344 (2007)
36. JY Cheah, JM Smith, Generalized M/G/C/C state dependent queueing models and pedestrian traffic flows. *Queueing Syst.* **15**(1), 365–386 (1994)
37. D Lee, S Zhou, X Zhong, Z Niu, X Zhou, H Zhang, Spatial modeling of the traffic density in cellular networks. *IEEE Wirel. Commun.* **21**(1), 80–88 (2014)
38. H Klessig, M Soszka, G Fettweis, in *2015 International Symposium on Wireless Communication Systems (ISWCS)*. Multi-cell flow-level performance of traffic-adaptive beamforming under realistic spatial traffic conditions, (2015), pp. 726–730

**Submit your manuscript to a SpringerOpen<sup>®</sup> journal and benefit from:**

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

---

Submit your next manuscript at ► [springeropen.com](http://springeropen.com)