Multiplexing Gain and Processing Savings of 5G Radio-Access-Network Functional Splits

Mohamed Shehata, Ahmed Elbanna, Francesco Musumeci, *Member, IEEE,* and Massimo Tornatore, *Senior Member, IEEE*

Abstract—Centralized Radio Access Network (C-RAN) is a promising mobile network architecture designed to support the requirements of future 5G mobile networks. In C-RAN, the "centralization" of baseband units enables substantial savings of computational resources (what we call "multiplexing gain" in this paper) and significant power savings. On the other hand, the deployment of C-RAN requires high capacity and imposes strict latency requirements on the fronthaul transport-network. To address these issues, various alternative architectures, known as "RAN functional splits", have been introduced to relax these strict fronthaul requirements. In this paper, we perform a quantitative analysis of the computational savings and the resulting power savings enabled by C-RAN, considering different RAN functional splits. To this end, we analytically model RAN computational resources to evaluate the multiplexing gain for different RAN functional splits. This model allows to calculate the processing reduction occurring in each RAN functional split. We then use this model to estimate the power savings of the various functional splits, considering different assumptions in terms of geographical areas, users distribution and number of aggregated cell sites. We find that up to 28% computational resources savings and 24% power savings can be achieved through functional splits in comparison to Distributed RAN.

Index Terms—C-RAN, multiplexing gain, computational effort, functional splits, GOPS.

I. INTRODUCTION

To satisfy growing users' demands, operators are forced to continuously increase their mobile network capacity. Maintaining the current network architecture will lead to an unsustainable network-cost increase as well as to a dramatic expansion in the network power consumption [2]. Hence, minimization of network cost and energy consumption have become a necessity for mobile network operators. Moreover, in the context of impending 5G communications, the design and operation of the radio access network is expected to be challenged by very stringent constraints in terms of tolerable latency and required data rate [3], e.g., with end-to-end latency requirements below 1 ms for future 5G services such as tactile Internet [4]. In addition to this, 5G networks will have to support more than 250 Gb/s/km^2 in dense-urban areas, with devices' density in the order of several hundreds -or even thousands- per km^2 [5].

In a traditional Distributed Radio Access Network (D-RAN), the Base Station (BS) comprises two modules, (*i*) the

Mohamed Shehata, Francesco Musumeci and Massimo Tornatore are with the Department of Electronics, Information and Bioengineering, Politecnico di Milano, Italy, e-mail: (mohamedkhaled.shehata@polimi.it, francesco.musumeci@polimi.it, massimo.tornatore@polimi.it).

Ahmed Elbanna is with Scientific Institute for Infrastructure and Communication Services Consult (WIK), Germany, e-mail: (a.elbanna@wik.org).

A preliminary version of this paper has been published in [1].

Remote Radio Head (RRH) for transmission and reception of radio signals, Digital-to-Analog/Analog-to-Digital Conversion (DAC/ADC) of the baseband signals, frequency up/down-conversion and power amplification, and (*ii*) the Baseband Unit (BBU) performing the digital processing functions of layer 1, 2 and 3 (L1, L2, L3). As shown in Fig. 1(a) every BS hosts its "local BBU" and has a dedicated housing facility, which is not shared with other BSs. Hence, in D-RAN, power consumption as well as investment and maintenance costs increase linearly with the number of BSs. Given the rapid traffic growth envisioned for the next years, simply increasing BSs density in D-RAN does not represent a scalable solution. A novel network architecture, called Centralized Radio Access Network (C-RAN), has been proposed as a more scalable alternative to D-RAN in terms of both power and cost efficiency [2]. The main idea of C-RAN is that multiple BBUs are placed in a single physical location (BBU pool), which is connected to several RRHs through a high capacity "fronthaul" network, as shown in Fig.1(b). Thanks to this centralization, the baseband resources in the BBU pool can also be virtualized and shared among several BSs, and significant reduction in the overall computational resources can be achieved due to multiplexing gain [1]. BBU centralization also allows to share maintenance costs and power consumption among several BSs, and promotes the utilization of advanced interference-cancellation techniques such as the Coordinated Multipoint (CoMP) [6].

On the other hand, despite the advantages of C-RAN, the fronthaul network must be able to support very high bit rates with very low latency [7], leading to high transport network cost. This has motivated researchers to investigate an compromise solution that mitigates the fronthaul requirements while enjoying the centralization benefits, through a flexible distribution of the processing functionalities. This class of solutions is referred to as "RAN functional splits" and consists in splitting the processing functionalities between RRHs (that are referred as, Radio Units, RUs, in the context of functional splits) and BBU pool (that are referred as, Digital Unit, DUs, in the context of functional splits). Based on the multiplexing gain model we developed in [1], the objective of this paper is to quantify the multiplexing gain, and the resulting power savings, obtained by adopting different RAN functional splits. To this end, we first provide an analytical model that quantifies the amount of processing operations required by each of the elementary functions that need to be implemented in RANs. Then, to calculate the computational savings, we provide a model to estimate the multiplexing gain that arises in the DUs
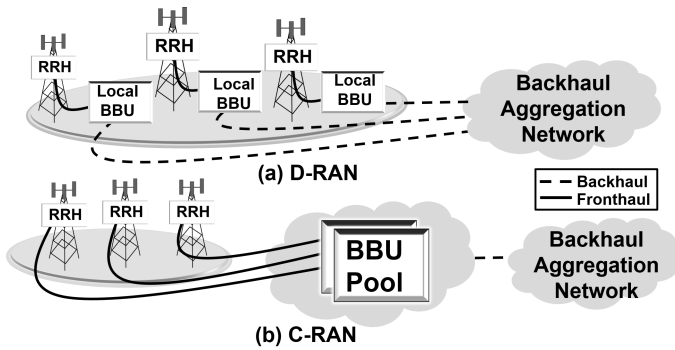
Fig. 1: Distributed RAN VS Centralized RAN.

under four different functional RAN split architectures. Our multiplexing-gain model takes into account several factors, namely: i) the different geographical areas that can be served by cell sites (geotypes), ii) spatial distribution of users, iii) users-eNB association strategy, iv) the scheduling algorithm adopted for the eNBs' physical resource blocks assignment to users. Finally, to estimate the power savings -resulting from reduction in the computational resources- for the various cases, we identify the main power consumption contributors in a BS and provide a power consumption model for the different RAN split options.

The remainder of this paper is organized as follows. Section II discusses related work regarding RAN functional splits, mobile-network power consumption and multiplexing-gain modeling. Section III discusses the local BBU architecture, the BBU pool architecture and the four considered RAN functional split architectures. Section IV presents a detailed model to capture the multiplexing gain for these four splits. Section V introduces our power modeling approach, based on the multiplexing gain model developed in the previous section. Section VI shows the illustrative numerical results. Section VII concludes the work.

II. RELATED WORK

Several works have been conducted on functional splits. In [8] the authors discuss the different functional-splits requirements in terms of capacity and latency, and show how different functional splits could be implemented according to the transport-network characteristics. Similar studies can be found in [9], [10] and [11]. In [12], the authors present a model to calculate the fronthaul bandwidth and the computational resources required for different functional RAN splits. In [13], authors present a comparison of the achievable pooling and CoMP gains for different RAN splits. Ref. [14] quantifies experimentally the transport-network capacity requirements of three different RAN splits, in terms of control plane and user plane data. Ref. [15] proposes a RAN split architecture called split-PHY that reduces the fronthaul bandwidth while keeping CoMP transmission and reception performance close to the C-RAN solution. Ref. [16] proposes a genetic algorithm to properly split and place the baseband functions with the objective of reducing the transport network cost. Note that none of the aforementioned works specifies the exact baseband

functions implemented at the RUs side and the DUs side for each split case.

Energy consumption of mobile networks has been subject of intensive research. Ref. [17] estimates the daily energy consumption of a 5G radio access technology (denoted as 5G-NX). The authors consider a typical European country, and reveal that, by using 5G technology, 55% of energy savings can be achieved while providing up to 15 times more capacity and 9 times higher peak rate compared to LTE network. Ref. [18] identifies a significant power saving by introducing two different downlink transmission strategies (namely, data-sharing and compression strategy) in C-RAN. Ref. [19] models the power savings achieved by applying dynamic BBU-RRH mapping, showing that, compared to D-RAN, 70% power savings can be achieved. Ref. [20] evaluates the power savings achieved in C-RAN by applying a cooperative transmission scheme with low computational complexity to mitigate interference. In [21], authors propose and solve a Virtual BS Formation (VF) optimization problem in C-RAN, and quantify the energy savings achieved by C-RAN with VF compared to D-RAN and C-RAN without VF.

Multiplexing gain that arises from the aggregation of multiple BBUs in a central pool has also been investigated. Ref. [22] presents a multi-dimensional Markov model to derive the multiplexing gain. Ref. [23] estimates the multiplexing gain based on a simulation study in the case of multiple sectors aggregated into a single cloud BS. Ref. [24] presents traffic simulation experiments to evaluate the multiplexing gain in WiMAX BSs under different traffic conditions. In [25], based on realistic data profile, the authors show that the centralized architecture can save at least 22% in computational resources compared to a distributed architecture by taking advantage of the variations in traffic and processing loads among the BSs. A similar study in [26] shows that reduction in the computation resources is up to 70%. Ref. [27] proposes a model to analyze the fronthaul statistical multiplexing gain brought by the spatial randomness of the traffic when aggregating several remote radio units as a cluster to share a fronthaul link.

To date, very few works considering power consumption and multiplexing gain of different RAN functional splits have appeared. Ref. [28] discusses the multiplexing gain in different RAN splits considering different traffic models. The author show the trade-off between the multiplexing gain and the cost of transport network. Ref. [29] presents a mixed-integer-constraint optimization problem to optimally split the baseband functions between the RUs and the DUs while minimizing the network power consumption and the transport network bandwidth.

So far, no study has evaluated the power consumption of the different RAN functional splits considering the multiplexing gain arising in each split option. To the best of our knowledge this is the first work that discusses the computational and power savings arising from different RAN functional splits and that specifies the exact baseband functionalities implemented at the RUs side and at the DUs side in each split.
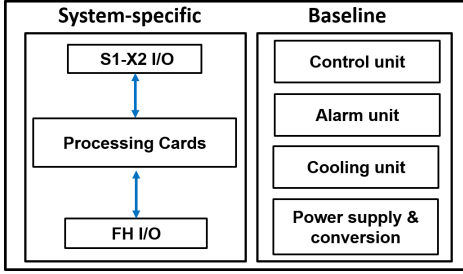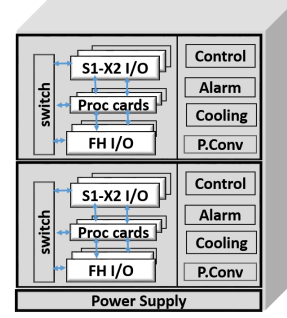
Fig. 2: Local BBU architecture.

III. OVERVIEW OF CENTRALIZED RADIO ACCESS NETWORK ARCHITECTURE

A. Local BBU and C-RAN BBU Pool

In this subsection we discuss our models for local BBU and BBU pools. The schematic model of a local BBU architecture (taken from [30]) is shown in Fig. 2. The local BBU is composed by two main parts, namely a baseline unit and system-specific unit. The baseline unit comprises components for control, alarms, cooling or fans, and power supply, while the system-specific unit includes the backhaul interface (S1-X2), the fronthaul interface, (e.g., Common Public Radio Interface (CPRI) [31]), and the processing cards.

The architecture of a BBU pool is shown in Fig. 3. Here, BBUs share some portions of their hardware resources. The rack in the example of Fig. 3 comprises two shelves, each one containing multiple processing cards. Processing cards are interconnected by a low-latency switch. Note that the amount of processing cards compared to the D-RAN case is reduced due to the arising multiplexing gain (we will elaborate on the meaning of multiplexing gain in Section IV). Finally, a baseline unit is mounted per shelf, while the main power supply is shared through the rack.

B. Functional RAN Splits

Mobile operators are seeking new solutions for a more flexible distribution of baseband functionalities between RUs and DUs [32]. Based on this more flexible distribution (i.e., the "RAN functional splits") some functionalities of the 3GPP LTE RAN protocol stack are executed at the RUs and others at the DUs.

In principle the separation of the functions (or in other words, the "splitting of the protocol stack") can be applied on any protocol layer, or on the interfaces between layers. To clarify how this splitting is performed, before introducing the different RAN functional splits we overview the various functions, grouped according to the protocol layer, as defined in 3GPP LTE RAN [33].

(i) **Physical layer (PHY layer)** is responsible for preparing the bit stream for transmission by executing some baseband functionalities:

- **Filtering** limits the signal bandwidth using lowpass filter (LPF), and then **sampling** the signal at Nyquist sampling rate.

Fig. 3: BBU pool architecture.

- **Fast Fourier Transform (FFT)** is a digital signal processing technique by which the sampled symbols are transferred to the frequency domain.
- **Resource demapping** allocates the subframes to their subcarriers.
- **Channel estimation** estimates the channel state information (CSI) using the pilot reference symbols in the received signal.
- **Predistortion** drives the power amplifier to work in the linear operating region.
- **MIMO precoding** constructs the spatial mapping matrix using the CSI of the users.
- **Equalization** compensates the effects of interchannel interference in a multipath fading channel.
- **OFDM demodulation** represents the binary data stream of the users with one of the following schemes: BPSK, QPSK 16QAM and 64 QAM.

(ii) **Medium Access Control (MAC) layer** is responsible for the channel coding through hybrid automatic repeat request (HARQ) and connects the radio link control layer to the PHY layer.

(iii) **Radio Link Control (RLC) layer** implements channel coding through automatic repeat request (ARQ). Also it implements the time-domain estimation/compensation of non-idealities which occur due to carrier frequency offset and sampling frequency offset.

(iv) **Packet Data Convergence Protocol (PDCP) layer** performs ciphering, integrity protection and IP header compression.

Fig. 4 shows the four possible functional splits considered in this paper.

CPRI split: Fig. 4(a) shows the CPRI split, where all baseband functionalities are located at the DUs (in CPRI split we can refer to the DUs as the BBU pool), while only power amplification and radio-frequency processing remain decentralized at the cell site (RU). This split maintains all the advantages of C-RAN as it enables highest multiplexing gain and maximum reduction of the complexity at RRH. On the other hand, this architecture should meet very strict latency requirements (in the order of 0.25ms [10]) for physical layer processing.

PHY split: Fig. 4(b) depicts the PHY split option, which splits the physical-layer functions into two parts, lower and upper physical layer. For this split, the latency requirement
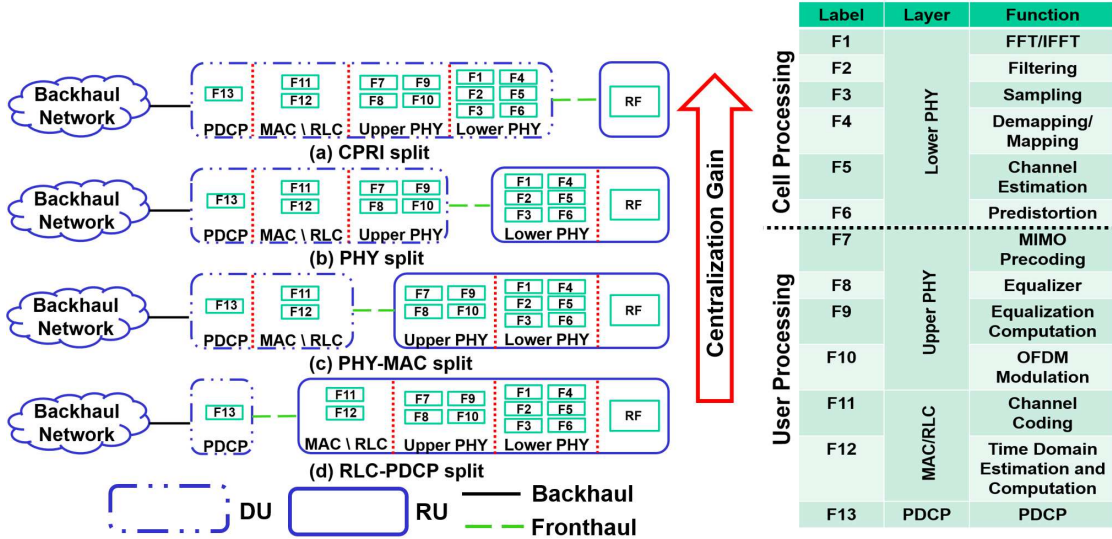
Fig. 4: RAN split options and corresponding mapping of network functions.

is relaxed to 2ms [10]. PHY split enables centralization of the upper PHY, MAC, RLC and PDCP functions, while the lower PHY functions (such as filtering, sampling, FFT/IFFT, resource mapping/demapping, and channel estimation as well as RF processing, A/D conversions and power pre-processing) are distributed.

PHY-MAC split: PHY-MAC split is shown in Fig. 4(c). The RAN architecture is split between the physical and the MAC layer. In this option, PHY layer functions are distributed, while, MAC, RLC and PDCP functions remain centralized at the DUs and impose 2ms latency requirement [10].

RLC-PDCP split: RLC-PDCP split is shown in Fig. 4(d). This split option applied the functional separation between the RLC and the PDCP layers. PDCP functions such as data packets header compression, ciphering, integrity remain centralized, while PHY, MAC and RLC functions are distributed. The centralized functions are not sensitive to latency (requirements are in the order of 30ms) [10] since all the scheduling functions (e.g., MIMO precoding and OFDM modulation) are distributed at the cell site.

IV. AN ANALYTICAL MODEL FOR MULTIPLEXING GAIN EVALUATION

In this section, we present an analytical model to evaluate the multiplexing gain as function of users distribution, users-eNBs association strategy and resource-blocks scheduling algorithm. First, we choose a statistical distribution (namely, normal or uniform) to model the spatial distribution of mobile users in a given serving area. Then we allocate the users to their serving eNBs according to a realistic user-eNB association strategy. After that, we apply a scheduling algorithm to distribute the physical resource blocks of the eNB among its users. Then, we calculate the computational effort (expressed in Giga Operations Per Second (GOPS)) per user after knowing channel condition, used resources, modulation, code rate and MIMO mode. Finally, starting from the statistical properties of the computational effort of the entire cell, we are

able to estimate the multiplexing gain. All the variables used in our model are summarized in Table I.

The various steps to build our model are detailed below.

User-eNB Association Strategy: We assume $|U|$ users are distributed (normally or uniformly) over an area covered by $|K|$ cell sites. Given transmitted power P_{T_k} of site k, distance d_{uk} between site k and user u, and a Rayleigh-distributed fading, the received power of user u can be defined as in [34], [35][1], i.e.:

$$P_{R_{uk}} = P_{T_k} - (128.1 + 37.6 \log (d_{uk})) \tag{1}$$

Then, user u will be allocated to site k^* with the highest received power where, $k^* = \arg \max_k P_{Ruk}$.

Scheduling Algorithm: We consider an LTE wireless system [36] in which the smallest scheduling allocation unit comprises two Physical Resource Blocks (PRBs), which are referred to as Scheduling Block (SB). Let us assume that each site k has S scheduling blocks with N_{SB} adjacent subcarriers each. Note that, in LTE, one Transmission Time Interval (TTI) equals the scheduling block duration (T_{SB}) and one frame consists of 10 TTIs. Let $Y = \{1, 2, \ldots n, \ldots N\}$ be the set of indices associated to the possible Modulation and Coding Schemes (MCS) in the system.

Consider user u uses the modulation scheme m_n with an associated code rate c_n over SB s, then the rate of user u over SB s can be calculated as follows [36]:

$$r_{us} = \frac{c_n \log_2 m_n}{T_{SB}} N_{SB} \tag{2}$$

The highest possible MCS which can be used by user u over SB s is determined if the Signal to Noise Ratio (SNR) of user u over SB s (SNR_{us}) is greater than a predefined SNR

---

[1]Note that the pathloss equation is given by: $L(dB) = 40(1 - 4 \times 10^{-3})log(d_{uk}) - 18log(Dhb) + 21log(f) + 80dB$ where d_{uk} is the distance between the user and cell site in kilometers, f is the carrier frequency in MHz and Dhb is the BS antenna height in meters. Considering a carrier frequency of 2000 MHz and a BS antenna height of 15 meters, the formula becomes $L(dB) = 128.1 + 37.6log(d_{uk})$.

TABLE I: Multiplexing Gain Model Variables.

Parameter	Description
U	Set of users where $u \in U$
K	Set of cell sites where $k \in K$
S	Set of scheduling blocks where $s \in S$
d_{uk}	Distance between user u and site k
P_{T_k}	Transmitted power of site k
$P_{R_{uk}}$	Power received by user u from site k
N_{SB}	Number of subcarriers per scheduling blocks
N	Index of the highest MCS in the system where $n \in N$
M_u	Index of the highest MCS could be used by user u
SNR_{us}	Signal to noise ratio of user u over SB s
r_{us}	Data rate of user u over SB s
m_n	Modulation scheme with index n
c_n	Code rate with index n
w_{usn}	set to 1 if SB s is allocated to user u using MCS n
T_{SB}	Scheduling block duration

threshold ($SNR_{Threshold}$) as stated in [37]. M_u is the index of the highest MCS that can be used by user u over all the SBs. For SNR_{us} calculation, Effective Exponential SNR Mapping (EESM) is used [38]. Then, the sets of allocated SBs and used MCSs for user u are obtained by maximizing the following quantity

$$\sum_{u=1}^{|U|} \sum_{s=1}^{|S|} \sum_{n=1}^{M_u} w_{usn} \frac{c_n \log_2 m_n}{T_{SB}} N_{SB} \qquad (3)$$

where w_{usn} is an assignment variable, which is set to 1 if SB s is allocated to user u using MCS n and set to 0 otherwise.

To perform the user-MCS association and the allocation of SB to users as stated in (3), we impose that a given user u can only use the same MCS over all his allocated SBs s in a given TTI, and a given SB s can only be allocated to one user in a given TTI [36]. Maximization of Eq. (3) is obtained as presented in [36]. MIMO mode and spatial streams are obtained knowing MCSs index [39]. Hence, in a given TTI, we know the SBs allocated to user u, used MCSs, MIMO mode and spatial streams.

A. Computational Effort and Multiplexing Gains for CPRI Split Option (C-RAN)

The computational effort for user u at TTI t can be defined as [23]:

$$CE(u,t) = \left(3A_{u,t} + A_{u,t}^2 + \frac{1}{3}M_{u,t}C_{u,t}L_{u,t}\right)\frac{R_{u,t}}{10} \qquad (4)$$

where $A_{u,t}$ is the number of used antennas for user u at TTI t, $M_{u,t}$ is the modulation bits for user u at TTI t ($\log_2 m_n$), $C_{u,t}$ is the code rate for user u at TTI t, $L_{u,t}$ is the number of spatial MIMO-layers for user u at TTI t, and $R_{u,t}$ is the number of SBs for user u at TTI t. CE is expressed in Giga Operations per Second (GOPS).

The computational effort for the whole network can now be calculated for both distributed RAN $CE_{DRAN_{total}}$ and centralized RAN $CE_{CRAN_{total}}$, by summing up the computational efforts for all the users in all the TTIs as follow

$$CE_{DRAN_{total}} = \sum_u \sum_t CE_{DRAN}(u,t) \qquad (5)$$

$$CE_{CRAN_{total}} = \sum_u \sum_t CE_{CRAN}(u,t) \qquad (6)$$

where $CE_{CRAN}(u,t)$ is computational effort computed by Eq. (4) for a user served by centralized RAN and $CE_{DRAN}(u,t)$ is computational effort computed by Eq. (4) for a user served by distributed RAN.

In a D-RAN, we calculate the computational effort separately for each cell. Each cell is expected to have a very different number of served users, which leads to a high variance in the computational effort of cells. In C-RAN, computational effort is calculated for the whole centralized pool resulting in lower variance in computational effort of the covered zone compared to the distributed one. Hence, in C-RAN it is more likely that the computational effort will not exceed a certain threshold with a certain probability p. The multiplexing gain β represents the percentage of savings in Computational Effort (CE) in the case of C-RAN with respect to the D-RAN case. CE is first calculated for both D-RAN (as in Eq. 7) and C-RAN (as in Eq. 6) for each TTI. Then, we draw two Cumulative Distribution Functions (CDF): 1- the first CDF for the CE values of the D-RAN 2- the second CDF for the CE values of the C-RAN. Finally, the multiplexing gain is calculated as the normalized difference between the CE of D-RAN and C-RAN at a given probability p as follows

$$\beta = \frac{CE_{DRAN_{total}} - CE_{CRAN_{total}}}{CE_{DRAN_{total}}} \qquad (7)$$

B. Computational Effort and Multiplexing Gains for Other RAN Functional splits

The calculation of computational effort for the other three functional RAN splits, i.e., PHY split, PHY-MAC, RLC-PDCP, differs from the CPRI split option seen above, as the overall computational effort at the DU is reduced. Note that, the highest multiplexing gain is achieved in the CPRI case, and it reduces gradually when fewer functions are centralized in the DUs. Hence, to model the multiplexing gain in the three different functional RAN splits (i.e., PHY, PHY-MAC, RLC-PDCP splits), we consider a scaling factor that represents the share of functions that are centralized in each specific split option with respect to the CPRI option. In other word, the user complexity is still calculated as in Eq. (4), but now a complexity factor σ is added to account only for the functions that are actually centralized. We calculate the complexity factor σ for each RAN split, based on the number of functions accommodated in the DUs as follows

$$\sigma = \frac{GOPS_{DU}}{GOPS_T} \qquad (8)$$

where $GOPS_{DU}$ is the GOPS of the baseband functionalities implemented in the DUs and $GOPS_T$ is the GOPS for the fully-centralized CPRI option.

For example, in PHY split , the user-processing functionalities are all implemented in the DUs; hence, by applying Eq. (8) we have a complexity factor σ of 0.6. In PHY-MAC split, where fewer functions are implemented in the DUs, complexity factor σ goes down to 0.3. In RLC-PDCP split, only core PDCP functions remain centralized at the baseband

pool and complexity factor σ equals to 0.1. In conclusion, the computational effort CE_{rs} for user u at TTI t in the RAN split rs can be defined as

$$CE_{rs}(u, t) = \sigma CE(u, t) \tag{9}$$

And, the multiplexing gain for the different RAN splits can be calculated again as in Eq. (7) but considering the computational effort as in Eq. (9).

V. ANALYTICAL MODEL FOR RAN POWER CONSUMPTION

A. Local BBU Power Consumption

As already mentioned in Section III, C-RAN introduces significant power savings, which come from the centralization of BBUs.

The power consumption of a local BBU[2] P_{BB} can be calculated as

$$P_{BB} = (P_{BL} + P_{FH} + P_{S1-X2} + P_{PC})(1 + \alpha_{PS}) \tag{10}$$

where, P_{BL} is the power consumption of baseline unit, P_{FH} is the power consumption of fronthaul interface, P_{X1-S2} is the power consumption of S1-X2 (backhaul) interface, P_{PC} is the power consumption of the processing card, α_{PS} is the power supply loss factor which is equivalent to the amount of dissipated power.

The power consumption of each of the mentioned components is evaluated as follows.

Processing Cards: The power consumption of the processing card depends on the complexity of the implemented functions (FFT, channel coding, modulation, etc) and on the processed traffic load. We use the power model in [40] to evaluate the power consumption of the processing cards.

This model is based on the complexity values, which estimate the number of GOPS for each baseband processing function. To calculate the power consumption of a baseband function P_i, the complexity value C_i of this baseband function must be converted into watts by using a technology dependent factor T, i.e.:

$$P_i = \frac{C_i}{T} \tag{11}$$

This technology-dependent factor indicates the hardware complexity according to the year of deployment and the system configuration. Values for these factors can be found in [41]. Note that, the complexity values of any baseband function are referred to a specific scenario (reference scenario), where bandwidth is 20 MHz, system is fully loaded (no hardware deactivation), antennas are single-input-single-output, and a spectral efficiency of 6 bps/Hz is achieved with a 64-QAM (Quadrature Amplitude Modulation) modulation scheme and a coding rate of 1.

For any other scenarios, P_i is scaled using factor Γ_i ($\Gamma_i = 1$ for reference scenario). This factor is determined according to the following parameters: 1) bandwidth, 2) spectral efficiency, 3) number of antennas, 4) system load in terms of hardware activation and deactivation (sleep-states), 5) number of spatial streams that are relevant to the number of antennas, and 6) the quantization (e.g., 4 bits, 16 bits and 24 bits).

Finally, the power consumption of the processing card is obtained through the summation of power consumed by each implemented baseband function as:

$$P_{PC} = \sum_i P_i \Gamma_i \tag{12}$$

A more detailed explanation of the model used for the processing cards power consumption can be found in [40].

Baseline Unit: The baseline unit usually comprises components responsible for system powering, power conversion (AC/DC and DC/DC), alarm unit, control unit, cooling unit (fans), and power supply. For simplicity, the baseline unit power is considered as a fixed value here, which linearly depends on the total power of the BS (power amplifier, analogue front-end, and baseband processing) [41]. Some efficiency factors for cooling, and power conversion (DC/DC, AC/DC) are introduced. Since we exclude the power amplifier and the analogue front end from our model, we consider the baseline unit power which linearly depends only on the baseband processing cards power as:

$$P_{BL} = P_{PC}\left(1 + \eta_{cool}\right)\left(1 + \eta_{dc/dc}\right)\left(1 + \eta_{ac/dc}\right) \tag{13}$$

where η_{cool} is an efficiency factor for the cooling unit, and $\eta_{dc/dc}$ & $\eta_{ac/dc}$ are the efficiency factors for DC/DC, AC/DC power conversions.

Fronthaul: In the different RAN split options, BSs are connected via fronthaul links towards the DUs. We assume a power consumption of a fronthaul interface $P_{FH} = 18.2$ W[3] [42].

Backhaul: Backhaul network performs traffic aggregation and transport between the RAN and the core network. In a typical macro BS, the backhaul interface power consumption is $P_{S1-X2} = 10$ W [40].

Power Supply: The efficiency of power supply is influenced by many factors. The exact efficiency numbers depend on the system configuration including system load and the year of deployment. For simplicity, we assume a fixed power loss α_{PS} = 5% in Local BBUs and α_{PS} = 10 % in different RAN splits [30].

B. DUs Power Consumption

The total power consumption P_T of a centralized digital unit (as depicted in Fig. 3) is calculated as:

$$P_T = (s(P_{BL} + P_{LS}) + s\nu(P_{FH} + (1-\beta)P_{PC}) + s\gamma P_{X1-S2})(1 + \alpha_{PS}) \tag{14}$$

where P_{Ls} is the power consumption of low latency switch, s is number of shelves, ν number of aggregated sites per shelf, β is multiplexing gain ($\beta < 1$) and γ is number of S1-X2 interfaces per shelf.

Moreover, the number of fronthaul interfaces per shelf is equal to ν the number of cells aggregated per shelf. The

---

[2]Note that, since analogue front-end and power amplification remain always at cell site, we exclude them from our analysis as they have no contributions in power savings.

[3]Although the required transport capacity depends on the considered RAN split, for simplicity we assume constant power consumption in the different functional RAN splits.
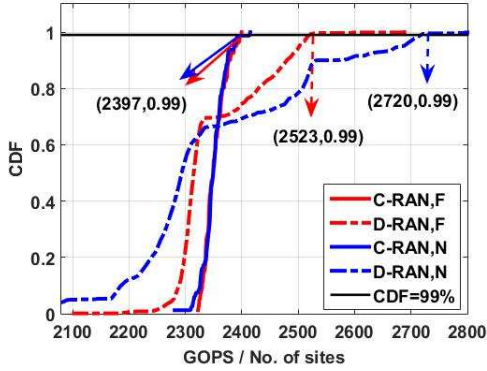
Fig. 5: Cumulative distribution function for computational effort in dense urban area considering normal and uniform users' distributions.

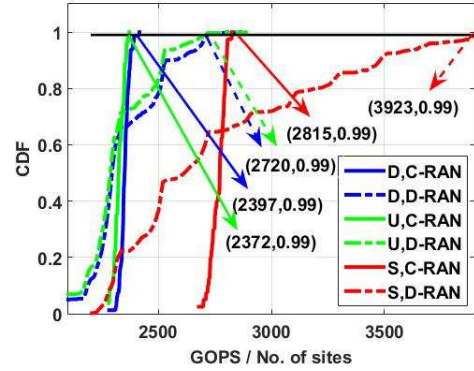
Fig. 6: Cumulative distribution function for computational effort considering three different geotypes with normal users' distribution.

number of processing cards per shelf equals the number of fronthaul interfaces per shelf multiplied by the multiplexing gain ($v(1 − \beta)$). The ratio between the number of S1-X2 interfaces per shelf to the number of processing cards per shelf is assumed to be 1:6, i.e., $\gamma = v(1 − \beta)/6$. The number of shelves is calculated as the minimum number of shelves that supports a given number of cells, i.e., Number of shelves $= \lceil$(No. of Cells)/(No. of Cells per Shelf)$\rceil$.

VI. ILLUSTRATIVE NUMERICAL RESULTS

In this section, we compare multiplexing gain and power consumption for a C-RAN considering the different functional RAN splits.

A. Evaluation Settings

We consider three different geographical type areas (geotypes), namely, Dense urban (D), Urban (U), and Sub-urban (S). The number of cell sites per unit area and number of users per unit area are taken from [42] and [43], respectively, and reported in Table II. In this study, we assume a particular scenario where a BS with 20 MHz bandwidth, 2x2 MIMO antenna configuration, 6 bps/Hz spectral efficiency, 64 QAM modulation scheme, coding rate of 1, and a full system load (no hardware deactivation).

B. Multiplexing Gain Assessment

We start observing the statistical properties of the computational effort calculated using Eq. (4) to calculate the

TABLE II: Features of the Considered Geotype.

	Dense Urban	Urban	Suburban
Number of Sites per Km^2	4	1.5	0.2
Total Area to Accommodate 12 site [km^2]	3	8	60
Number of User per km^2	3000	1000	500
Total Number of Users	9000	8000	30000

multiplexing gain and show the impact of the users distribution, different geotypes and the different pool dimension, for different functional RAN splits.

Fig. 5 shows the cumulative distribution function (CDF) of the computational effort (CE) in a dense-urban area, where users are distributed normally (N) and uniformly (F) for the case of C-RAN (CPRI split) and the case of D-RAN. We run the simulation for 80 TTIs, and a different GOPS value in each TTI is obtained. So, we draw a graph where y-axis represents the cumulative probability of achieving a certain number of GOPS in a TTI averaging over 80 trials (80 TTIs). We calculate the multiplexing gain as follows. We assume a network operator designs a baseband pool with enough CPUs to serve the requested amount of GOPS with 99% probability (in other words, we only admit blocking of 1% of requested computational effort). So, multiplexing gain can be calculated as the percentage difference between the amount of GOPS required to achieve 99% of the CDF in the case of one single (distributed) eNB (D-RAN) and when aggregating 12 eNBs (C-RAN). In Fig. 5 for example, multiplexing gain for normal distribution is calculated as follows

$$\beta = \frac{2720 − 2397}{2720} \times 100 = 11.8\% \qquad (15)$$

while multiplexing gain for uniform distribution is calculated as

$$\beta = \frac{2523 − 2397}{2523} \times 100 = 5\% \qquad (16)$$

The multiplexing gain is lower in case of uniform distribution. In fact, when users are normally distributed, they are more concentrated closer to the eNBs and this results in higher computational effort with respect to the uniform case, which leads to higher multiplexing gain.

Fig. 6 compares the computational effort in the different geotypes (considering CPRI split, 12 eNBs and normally distributed users). Surprisingly, sub-urban area has the maximum computational effort while the dense urban has the lowest one. This can be logically explained if we consider that users' SNR in sub-urban area is higher since the number of users per unit area is lower, hence; users can more likely transmit with high air-interface configuration, and hence, the computational effort
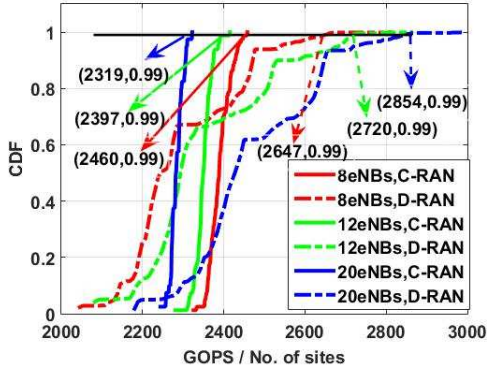
Fig. 7: Cumulative distribution function for computational effort in dense urban area considering different pool dimensions with normal users' distribution.

Fig. 8: Multiplexing gain for different RAN splits considering different geotypes with normal users' distribution.

per user is higher with respect to urban and dense urban. Moreover, a large number of SBs will be occupied in sub-urban area as the total number of users is high[4] leading to highest computational effort. On the contrary, the dense urban area has the lowest computational effort. Accordingly, sub-urban area has the maximum multiplexing gain while the dense urban has the lowest one. In summary, gain is 11.8% in dense urban area, 13% in urban area, and 28.2% in sub-urban area. Note that our comparison is referred to a fixed number of aggregated eNBs per pool, which is constant for all the considered geotypes. This might not be necessarily the case and depends on mobile operators planning choices. Note also that, even if multiplexing gains are the highest in the suburban area, the high cost of the fronthaul network to cover such a large area might anyway discourage operators from adopting C-RAN in suburban areas (considerations regarding cost of the fronthaul are out of scope in this paper).

Fig. 7 shows that increasing the pool dimensions, i.e. the number of aggregated eNBs, for a given geotype area results in higher multiplexing gain. We consider CPRI split option and a dense urban area with different pool dimensions: 8 eNBs, 12eNBs and 20s eNB per pool. The obtained gain in the case of aggregating 20 eNBs is 19%, 11.8% when aggregating 12 eNBs and 7% when aggregating 8 eNB.

Finally, in order to capture the multiplexing gains obtained for different RAN functional splits in the different geotypes, we apply Eq. (7) and Eq. (9) considering 12 eNBs with normally distributed users. The results shown in Fig. 8 quantify the decrease of the multiplexing gain due to the adoption of less aggressive functional splits.

C. Power Consumption Assessment

This subsection compares the BBU pool power consumption estimated using the power model in Section V, for different functional RAN splits. We show the power savings in each split option compared to D-RAN. We refer to CPRI split, PHY split, PHY-MAC split and RLC-PDCP split as S1, S2, S3 and S4, respectively.

<hr>

[4]Note that the area dimension in sub-urban is very high to accommodate 12 eNBs as in Table II.
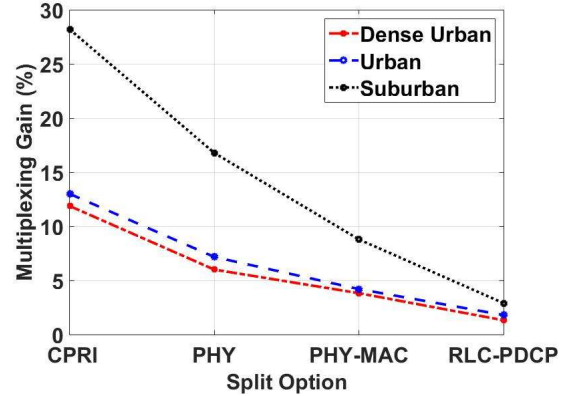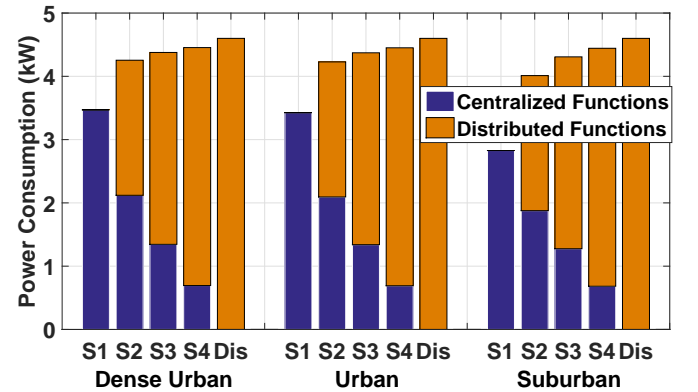

Fig. 9: Power consumption for 12 aggregated cell sites considering different RAN splits in three different geotypes with normal users' distribution.

Fig. 9 shows the power consumption of 12 aggregated cell sites considering the different functional RAN splits and full distributed solution in dense-urban, urban, and sub-urban scenarios. In dense urban scenario, we obtain a power savings of 24% with respect to distributed RAN (Dis) in CPRI split, a 7.6% in PHY split, a 4.8% in PHY-MAC split, and 3.1% in RLC-PDCD split. In urban scenario, we obtain a power saving of 25.5% with respect to distributed RAN in CPRI split, a 8.1% in PHY split, a 5% in PHY-MAC split, and 3.2% in RLC-PDCD split. In sub-urban scenario, we obtain a power savings of 38.5% with respect to distributed RAN in CPRI split, a 12.7% in PHY split, a 6.35% in PHY-MAC split, and 3.4% in RLC-PDCD split. Note that the more centralized functionalities in the baseband pool the less total power consumed, due to the fact that the higher consolidation the less computational resources are required, leading to higher power savings.

In Fig. 10 we show power consumption of 20 aggregated cell sites considering four different functional RAN splits in dense-urban scenario. Compared to distributed RAN, we obtain a power saving of 12% in PHY split, 9.2% in PHY-MAC split, and 7.32% in RLC-PDCD split. As expected power savings increase when centralizing more cells.
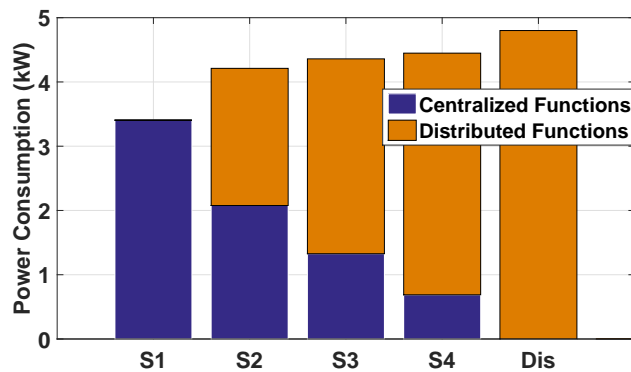
8

Fig. 10: Power consumption for 20 aggregated cell sites considering different RAN splits in dense urban area with normal users' distribution.

VII. CONCLUSION

In this work, we propose a multiplexing gain model to capture the processing savings arising from consolidation of compute resources, considering four different functional RAN splits: CRPI, PHY, MAC-PHY and RLC-PDCP. We also propose a power consumption model to quantify the savings associated with each split option compared to the traditional one (D-RAN).

Results prove that the highest multiplexing gains (accordingly, the lowest power consumptions) are obtained for CPRI split in all different geotypes, and show how multiplexing gain decreases significantly when adopting less aggressive splits, as RLC-PDCP split. For 12 aggregated sites in dense urban area, we estimate savings up to 11.8% in CPRI, 6% in PHY, 3.8% in MAC-PHY and 1.32% in RLC-PDCP. Under our assumptions, we found that the highest multiplexing gain is obtained in a suburban scenario. We also show that a higher multiplexing gain is achieved as the number of aggregated cell sites increases for a given geotype.

ACKNOWLEDGMENT

REFERENCES

[1] M. Shehata, A. Elbanna, F. Musumeci, and M. Tornatore, "C-RAN baseband pooling: Cost model and multiplexing gain analysis," in *Proc. 19th International Conference on Transparent Optical Networks (ICTON)*, Girona, Spain, 2017, pp. 1–4.

[2] A. Checko, H. L. Christiansen, Y. Yan, L. Scolari, G. Kardaras, M. S. Berger, and L. Dittmann, "Cloud RAN for mobile networks- A technology overview," *IEEE Communications surveys & tutorials*, vol. 17, no. 1, pp. 405–426, Mar. 2015.

[3] E. Hossain and M. Hasan, "5G cellular: key enabling technologies and research challenges," *IEEE Instrumentation & Measurement Magazine*, vol. 18, no. 3, pp. 11–21, Jun. 2015.

[4] M. Simsek, A. Aijaz, M. Dohler, J. Sachs, and G. Fettweis, "5G-Enabled Tactile Internet," *Journal of Selected Areas in Communication*, vol. 34, no. 3, pp. 460–473, Mar. 2016.

[5] "5G White Paper," Next generation Mobile Network (NGMN) Alliance, White paper, Feb. 2015. [Online]. Available: https://www.ngmn.org

[6] Y. Huiyu, Z. Naizheng, Y. Yuyu, and P. Skov, "Performance evaluation of coordinated multipoint reception in C-RAN under LTE-Advanced uplink," in *Proc. 7th International ICST Conference on Communications and Networking in China (CHINACOM)*, Aug. 2012, pp. 778–783.

[7] A. Pizzinat, P. Chanclou, F. Saliou, and T. Diallo, "Things you should know about fronthaul," *Journal of Lightwave Technology*, vol. 33, no. 5, pp. 1077–1083, Mar. 2015.

[8] A. Maeder, M. Lalam, A. De Domenico, E. Pateromichelakis, D. Wubben, J. Bartelt, R. Fritzsche, and P. Rost, "Towards a flexible functional split for cloud-RAN networks," in *Proc. European Conference on Networks and Communications (EuCNC)*, Bologna, Italy, Jun. 2014, pp. 1–5.

[9] J. Bartelt, P. Rost, D. Wubben, J. Lessmann, B. Melis, and G. Fettweis, "Fronthaul and backhaul requirements of flexibly centralized radio access networks," *IEEE Wireless Communications*, vol. 22, no. 5, pp. 105–111, Oct. 2015.

[10] Small Cell Forum, "Small Cell Virtualization: Functional Splits and Use Cases," Jan. 2016. [Online]. Available: https://www.scf.io

[11] M. Jaber, D. Owens, M. A. Imran, R. Tafazolli, and A. Tukmanov, "A joint backhaul and RAN perspective on the benefits of centralised RAN functions," in *Proc. IEEE International Conference on Communications Workshops (ICC)*, Kuala Lumpur, Malaysia, May 2016, pp. 226–231.

[12] X. Wang, L. Wang, S. E. Elayoubi, A. Conte, B. Mukherjee, and C. Cavdar, "Centralize or distribute? A techno-economic study to design a low-cost cloud radio access network," in *Proc. IEEE International Conference on Communications (ICC)*, Paris, France, May 2017, pp. 1–7.

[13] U. Dötsch, M. Doll, H.-P. Mayer, F. Schaich, J. Segel, and P. Sehier, "Quantitative analysis of split base station processing and determination of advantageous architectures for LTE," *Bell Labs Technical Journal*, vol. 18, no. 1, pp. 105–128, Jun. 2013.

[14] L. Valcarenghi, K. Kondepu, F. Giannone, and P. Castoldi, "Requirements for 5G fronthaul," in *Proc. 18th International Conference on Transparent Optical Networks (ICTON)*, Trento, Italy, Jul. 2016, pp. 1–5.

[15] K. Miyamoto, S. Kuwano, J. Terada, and A. Otaka, "Split-PHY processing architecture to realize base station coordination and transmission bandwidth reduction in mobile fronthaul," in *Proc. Optical Fiber Communications Conference and Exhibition (OFC)*, Los Angeles, CA, USA, Mar. 2015, pp. 1–3.

[16] J. Liu, S. Zhou, J. Gong, Z. Niu, and S. Xu, "Graph-based framework for flexible baseband function splitting and placement in C-RAN," in *Proc. IEEE International Conference on Communications (ICC)*, London, UK, Jun. 2015, pp. 1958–1963.

[17] S. Tombaz, P. Frenger, M. Olsson, and A. Nilsson, "Energy performance of 5G-NX radio access at country level," in *Proc. IEEE 12th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*, New York, USA, Oct. 2016, pp. 1–6.

[18] B. Dai and W. Yu, "Energy efficiency of downlink transmission strategies for cloud radio access networks," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 4, pp. 1037–1050, Apr. 2016.

[19] M. Khan, R. S. Alhumaima, and H. S. Al-Raweshidy, "Reducing energy consumption by dynamic resource allocation in C-RAN," in *Proc. European Conference on Networks and Communications (EuCNC)*, Paris, France, Jun. 2015, pp. 169–174.

[20] L. Chen, H. Jin, H. Li, J.-B. Seo, Q. Guo, and V. Leung, "An energy efficient implementation of C-RAN in HetNet," in *Proc. IEEE 80th Vehicular Technology Conference (VTC2014-Fall)*, Vancouver, BC, Canada, Sep. 2014, pp. 1–5.

[21] X. Wang, S. Thota, M. Tornatore, H. S. Chung, H. H. Lee, S. Park, and B. Mukherjee, "Energy-efficient virtual base station formation in optical-access-enabled cloud-RAN," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 5, pp. 1130–1139, May 2016.

[22] J. Liu, S. Zhou, J. Gong, Z. Niu, and S. Xu, "Statistical multiplexing gain analysis of heterogeneous virtual base station pools in cloud radio access networks," *IEEE Transactions on Wireless Communications*, vol. 15, no. 8, pp. 5681–5694, Aug 2016.

[23] T. Werthmann, H. Grob-Lipski, and M. Proebster, "Multiplexing gains achieved in pools of baseband computation units in 4G cellular networks," in *Proc. IEEE 24th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, London, UK, Sep. 2013, pp. 3328–3333.

[24] M. Madhavan, P. Gupta, and M. Chetlur, "Quantifying multiplexing gains in a wireless network cloud," in *Proc. IEEE International Conference on Communications (ICC)*, Ottawa, ON, Canada, Jun. 2012, pp. 3212–3216.

[25] S. Bhaumik, S. P. Chandrabose, M. K. Jataprolu, G. Kumar, A. Muralidhar, P. Polakos, V. Srinivasan, and T. Woo, "CloudIQ: A framework for processing base stations in a data center," in *Proc. 18th annual international conference on Mobile computing and networking (Mobicom '12)*, Istanbul, Turkey, Aug. 2012, pp. 125–136.

[26] S. Namba, T. Matsunaka, T. Warabino, S. Kaneko, and Y. Kishi, "Colony-RAN architecture for future cellular network," in *Proc. Future Network & Mobile Summit (FutureNetw)*, Berlin, Germany, Jul. 2012, pp. 1–8.

[27] L. Wang and S. Zhou, "On the fronthaul statistical multiplexing gain," *IEEE Communications Letters*, vol. 21, no. 5, pp. 1099–1102, May 2017.

[28] A. Checko, A. P. Avramova, M. S. Berger, and H. L. Christiansen, "Evaluating C-RAN fronthaul functional splits in terms of network level energy and cost savings," *Journal of Communications and Networks*, vol. 18, no. 2, pp. 162–172, Apr. 2016.

[29] X. Wang, A. Alabbasi, and C. Cavdar, "Interplay of energy and bandwidth consumption in C-RAN with optimal function split," in *Proc. IEEE International Conference on Communications (ICC)*, Paris, France, May 2017, pp. 1–6.

[30] M. De Andrade, M. Tornatore, A. Pattavina, A. Hamidian, and K. Grobe, "Cost models for BaseBand Unit (BBU) hotelling: From local to cloud," in *Proc. IEEE 4th International Conference on Cloud Networking (CloudNet)*, Niagara Falls, ON, Canada, Canada, Oct. 2015, pp. 201–204.

[31] *Common Public Radio Interface (CPRI); Interface Specification*, 2009. [Online]. Available: http://www. cpri. info

[32] C.-Y. Chang, N. Nikaein, and T. Spyropoulos, "Impact of packetization and scheduling on C-RAN fronthaul performance," in *Poc. IEEE Global Communications Conference (GLOBECOM)*, Washington, DC, USA, Dec. 2016, pp. 1–7.

[33] "LTE; Evolved Universal Terrestrial Radio Access (E-UTRA); Long Term Evolution (LTE) physical layer; General description," 3rd Generation Partnership Project (3GPP), Technical Specification (TS) 36.201, Apr. 2009, version 3.3.0. [Online]. Available: https://www.3gpp.org

[34] "LTE; Evolved Universal Terrestrial Radio Access (E-UTRA); Radio Frequency (RF) requirements for LTE Pico Node B," 3rd Generation Partnership Project (3GPP), Technical Specification (TS) 36.931, May 2011, version 9.0.0. [Online]. Available: https://www.3gpp.org

[35] "Universal Mobile Telecommunications System (UMTS); Radio Frequency (RF) system scenarios," 3rd Generation Partnership Project (3GPP), Technical Report (TR) 25.942, Apr. 2017, version 14.0.0. [Online]. Available: https://www.3gpp.org

[36] R. Kwan, C. Leung, and J. Zhang, "Proportional fair multiuser scheduling in LTE," *IEEE Signal Processing Letters*, vol. 16, no. 6, pp. 461–464, Jun. 2009.

[37] D. Lopez-Perez, A. Ladanyi, A. Jüttner, H. Rivano, and J. Zhang, "Optimization method for the joint allocation of modulation schemes, coding rates, resource blocks and power in self-organizing LTE networks," in *Proc. IEEE International Conference on Computer Communications (IEEE INFOCOM 2011)*, Shanghai, China, Apr. 2011, pp. 111–115.

[38] S. Mumtaz, A. Gamerio, and J. Rodriguez, "EESM for IEEE 802.16 e: WiMaX," in *Proc. 7th IEEE/ACIS International Conference on Computer and Information Science (ICIS)*, Portland, OR, USA, May 2008, pp. 361–366.

[39] R. Van Nee, V. Jones, G. Awater, A. Van Zelst, J. Gardner, and G. Steele, "The 802.11 n MIMO-OFDM standard for wireless LAN and beyond," *Wireless Personal Communications*, vol. 37, no. 3, pp. 445–453, May 2006.

[40] B. Debaillie, C. Desset, and F. Louagie, "A flexible and future-proof power model for cellular base stations," in *Proc. IEEE 81st Vehicular Technology Conference (VTC Spring)*, Glasgow, UK, May 2015, pp. 1–7.

[41] C. Desset, B. Debaillie, V. Giannini, A. Fehske, G. Auer, H. Holtkamp, W. Wajda, D. Sabella, F. Richter, M. J. Gonzalez *et al.*, "Flexible power modeling of LTE base stations," in *Proc. IEEE Wireless Communications and Networking Conference (WCNC)*, Shanghai, China, Apr. 2012, pp. 2858–2862.

[42] "Deliverable 3.3: Analysis of Transport Network Architectures for Structural Convergence," COnvergence of fixed and Mobile BrOadband access/aggregation networks- COMBO Project, Tech. Rep., Jul. 2015. [Online]. Available: https://www.ict-combo.eu

[43] G. Auer, V. Giannini, C. Desset, I. Godor, P. Skillermark, M. Olsson, M. A. Imran, D. Sabella, M. J. Gonzalez, O. Blume *et al.*, "How much energy is needed to run a wireless network?" *IEEE Wireless Communications*, vol. 18, no. 5, pp. 40–49, Oct. 2011.