Politecnico di Milano
Dipartimento di Elettronica, Informazione e Bioingegneria

DOTTORATO DI RICERCA IN INGEGNERIA
DELL'INFORMAZIONE

Energy Efficiency and Survivability in 5G Centralized Access Networks

Doctoral Dissertation of:
Mohamed Shehata

Advisor:
Prof. Massimo Tornatore

Tutor:
Prof. Matteo Cesana

Supervisor of the Doctoral Program:
Prof. Andrea Bonarini

2018 – Cycle XXXI

To my family

Abstract

The continuous demand for better wireless data services in terms of very high data rates (typically of Gbps order), extremely low latency, and significant improvement in users' perceived Quality-of-Service, has triggered the research on the fifth generation (5G) wireless systems that are expected to be deployed beyond 2020. 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. Hence, minimization of network cost and energy consumption have become a necessity for mobile network operators. In order to do so, the network infrastructure has to evolve from the old static architecture towards a more scalable, dynamic and agile one by resorting to novel technologies and architectural solutions to improve cost and energy efficiency.

Centralized Radio Access Network (C-RAN) is a promising mobile network architecture designed to support the requirements of future 5G mobile networks. C-RAN is a new mobile access network architecture where base stations are splitted among BaseBand Units (BBU) and Remote Radio Heads (RRH) and BBUs are centralized into single physical location, with the consequent introduction of the new "fronthaul" transport network. In C-RAN, the "centralization" of baseband units enables substantial savings of computational resources 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 thesis, we investigate the opportunities enabled by C-RAN. First we provide a quick survey on the C-RAN stat of art. Then, we model the computational savings (what we called multiplexing gain) enabled by C-RAN under four different functional splits. Furthermore, we show the cost savings arises from centralization. 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. Following this centralization savings, we design a survivable C-RAN against BBU pool and link failures. We propose three different approaches for the survivable BBU pool placement problem and traffic routing in C-RAN deployment

over a 5G optical aggregation network. We formalize different protection scenarios as Integer linear Programming (ILPs) problems. The ILPs objectives are to minimize the number of BBU pools, the number of used wavelengths and the baseband processing computational resources. Finally, we design survivable C-RAN based on shared path protection schemes with the objective to minimize the number of BBU pools and the number of used wavelengths.

The results show the cost and energy advantages of C-RAN with respect to classical RANs, due to "centralization" of BBUs into a few sites. Moreover, the results give insights on the latency and the transport network capacity on the BBU placement problem.

List of Publications

P1 **M. Shehata**, A. Elbanna, F. Musumeci and M. Tornatore. "C-RAN Baseband Pooling: Cost Model and Multiplexing Gain Analysis", in *proceedings of ICTON*, Girona- Spain, pp. 1-4, Jul. 2017.

The material of this publication contributes to Chapter 3.

P2 **M. Shehata**, A. Elbanna, F. Musumeci and M. Tornatore, "Multiplexing Gain and Processing Savings of 5G Radio Access Network Functional Split", *IEEE Transactions on Green Communications and Networking*, accepted for Publication Sep. 2018.

The material of this publication contributes to Chapters 3 and 4.

P3 **M. Shehata**, O. Ayoub, F. Musumeci and M. Tornatore, "Survivable BBU Placement for C-RAN over Optical Aggregation Networks", in *proceedings of ICTON*, Bucharest- Romania, pp. 1-4, Jul. 2018.

The material of this publication contributes to Chapter 5.

P4 **M. Shehata**, F. Musumeci and M. Tornatore, "Resilient BBU Placement in 5G C-RAN over Optical Aggregation Networks", submitted To *Photonic Network Communications (PNET), Springer*, Jul. 2018.

The material of this publication contributes to Chapter 5.

Contents

List of Figures

List of Tables

Introduction 1

1.1 Motivation

Over the past decades and since the first generation (1G) of mobile networks has been launched, Information and Communication Technologies (ICT) have witnessed a significant evolution. As a result, innovations with a real impact on the quality of life of the individuals and societies has been raised. These include smart cities applications, smart health, smart cars, Video on Demand (VoD) and extra high definition mobile applications. Moreover, there is a huge increase in the population of smart devices, according to Cisco and Ericsson's recent forecast, mobile portable devices and connections is anticipated to grow to 11.5 billion by 2019, and thereon to more than a ten-fold increase in mobile data traffic between 2013 and 2018 [1]. Additionally, in 2021, 150 billion devices will be 5G connected [2].

It is a challenge in the present Long Term Evolution (LTE) cellular system to support the huge and rapid increase of data usage and connectivity. Therefore, different features of LTE cellular networks have been introduced, such as, Multiple-Input Multiple-Output (MIMO) antennas, Coordinated Multi-Point (CoMP) transmission and Heterogeneous Networks (HetNets) to enhance capacity and data rates. However, it is unlikely to sustain this ongoing traffic explosion in the long run. As a result, the focus of the Mobile Network Operators (MNOs) is to accommodate the huge mobile data volumes.

In the present LTE cellular system, it has been a contradiction between fulfilling the huge data traffic demands and sustaining acceptable revenue margins. This contradiction is due to the fact that,

the reduction of the Total cost of Ownership (TCO) in the current network architecture is hard to be achieved by MNOs. Total Cost of Ownership (TCO) in mobile networks includes Capital Expenditure (CapEx) and Operational Expenditure (OpEx). CapEx mainly refers to expenditure relevant to network construction, Radio Frequency (RF) hardware, baseband hardware, installation, civil cost and site support. OpEx covers the cost needed to operate the network, i.e., site rental, energy consumption, operation and maintenance as well as site-upgrade. In fact, the explosion of data traffic volumes requires significant increase in the network capacity which leads to deploying more Base Stations (BSs). More specifically, CapEx increases as BSs are the most expensive components of a wireless network infrastructure, while OpEx increases as cell sites demand a considerable amount of energy to operate. For example, China Mobile estimates 72% of total power consumption originates from the cell sites [3]. MNOs need to cover the expenses for network construction, operation, maintenance and upgrade.

Energy efficiency in ICT brings more attention to both academic and industrial sectors for economical reasons -due to energy consumption significant contribution to the OpEx as mentioned previously- as well as for environmental reasons. Energy consumption of ICT forms 8% of the total worldwide energy consumption, and is anticipated to reach 14% by 2020 [4]. Therefore, this much needed energy is a significant contributor to carbon dioxide and Green House Gas (GHG) emissions. In the context of the awareness around climate change and its potentially devastating effects, reducing the energy consumption of ICT attracted attentions in recent years to mitigate the negative effects of global warming on our environment.

In order to overcome the above mentioned challenges (Traffic, Cost, Energy) ICT clearly needs new communications infrastructure the like of which we have never seen before. Therefore, novel architectures and technologies that optimize cost and energy consumption while satisfying the exploding data traffic volumes become a necessity in the field of mobile network. Whilst 5G promises major technical advances and better geographical coverage, it is very promising to be a solution for all the above challenges and requirements.

1.2 5G Requirements

The new fifth generation (5G) of mobile networks represents the future of mobile communications. As mentioned previously, the need for 5G

comes mainly from the exponentially increasing demand data traffic volumes, explosion of network cost and the continuous increase of energy consumption. It was promised that by 2020, the 5G wireless and mobile communication systems will have to provide about 1000 times more capacity than today, reducing up to 90% of the consumed energy per service [5]. 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[6].

5G exclusively introduces new services, which requires critical ultra-reliable low latency communication services (URLLC), and ultra-fast and reliable mobile connectivity. These new services are not only interconnecting people, but also objects, machines, and vehicles. 5G can help the industry achieve their goals by adjusting the mobile network to the required services. Therefore, a greater flexibility and service versatility will be two of the major improvements that will be introduced by 5G. According to the *Mobile and wireless communications Enablers for Twenty–Twenty Information Society* (METIS), 5G services can be categorized into three main services as follow [7]

- **Extreme Mobile BroadBand (xMBB)** also referred to as enhanced MBB (eMBB), provides increased data rates and low latency communication. xMBB includes peak rates in the order of Gbps and moderate rates, in the order of tens of Mbps with very high reliability. xMBB is considered to be a natural evolution to the data rates we enjoy today with currently deployed LTE. The ultimate 5G goal is to deliver up to 10-Gb/s peak throughput, 1-Gb/s throughput in high mobility, and up to 10,000X total network traffic. Meanwhile, technology developments such as millimeter waves and massive MIMO will play a critical role in achieving the eMBB goals.

- **Massive Machine-Type Communication (mMTC)** provides connectivity for up to tens of billions to cost and energy-constrained devices worldwide. The main feature of this service is the massive number of connected devices.

- **Ultra-reliable Machine-Type Communication (uMTC)** provides ultra-reliable, low latency and resilient connectivity for industrial control applications time-critical services, e.g., V2V (Vehicle-to-Vehicle). The main feature is high reliability with low number of connections and required data rate compared to mMTC.

In conclusion, the capabilities of 5G will extend far beyond those of the current LTE networks, therefore new technologies and architectures are needed. Those new technologies and architectures have to tackle the above mentioned cost and energy consumption challenges. Therefore, Centralized Radio Access Network (C-RAN) is introduced as one of the very promising solutions that optimizes cost and energy efficiency and will be discussed in the next section.

1.3 Centralized Radio Access Network

In a traditional Distributed Radio Access Network (D-RAN), the Base Station (BS) comprises two modules, (*i*) the 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). 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[8]. 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. 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[9]. 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)[10].

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[11], 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).

1.4 Survivable Networks

In the light of the previous mentioned rapid growth of the mobile users demands, wireless mobile networks become a part of everyday life. Therefore, the interruption of the service for even short periods of time may have fatal consequences in terms of quality of service and user satisfaction. In this context, how to prevent service failure and minimizing the failure time if occurred becomes a critical issue. Hence, "survivability" is one of the main requirements for mobile networks, which is the ability to provide and maintain an acceptable level of service in the face of various faults and challenges to normal operation [12].

In the context of C-RAN, it is an important aspect to deal with BBU pool and link failures, as these failures might cause service outage for a large area with a significant number of users. This has recently motivated novel research on the design of what know as "survivable C-RAN".

1.5 Contribution and Outline of the Thesis

The aim of this research is to investigate the cost and energy efficiency as well as the survivability of 5G C-RAN. Our research work starts with a survey of the technological features of C-RAN and implementation principles from a networking standpoint. Then an analysis of the energetic benefits that this technique can bring over currently aggregation infrastructure is performed. We design a model to capture the multiplexing gain that emanates from centralization. This model is able to estimate the computational savings in 5G C-RAN. Moreover, we explore the multiplexing gain for different RAN functional splits, showing the trade-off between the centralization gains and the fronthaul requirements. We use this model to calculate the economic aspects of C-RAN deployment and the cost reduction compared to the traditional distributed RAN. The research work is extended by proposing a model to calculate the power consumption in the different

RAN split options by identifying the main power consumption contributors. Finally, we design a survivable C-RAN to cope with link and node failures by evolving several ILP algorithms with multiple objective functions based on dedicated and shared protection. This thesis is divided into six chapters, including the current one, and they are structured as follows

- In **Chapter 2**, we focus on presenting an overview of the C-RAN, advantages, and challenges of its implementation are detailed. A classification of the various architectural solutions for BBU centralization is detailed.
- In **Chapter 3**, we present an analytical model to estimate the statistical multiplexing gain of a centralized baseband pool. Moreover, we perform a techno-economic study for the BBU pools deployment. Finally, we tackle the trade-off between the fronthaul requirements and the centralization gains. In this context, we investigate different RAN functional splits and extent our previous mentioned model to calculate the multiplexing gain for those RAN splits
- In **Chapter 4**, we focus on the power savings arising in C-RAN. We introduce a detailed power consumption model to estimate the power savings resulting from reduction in the computational resources discussed in the previous chapter. Our power model is built through identifying the main power consumption contributors in a BS namely: Baseband, Fronthaul interface, Backhaul interface, Baseline. Moreover, we extend our power model to estimate the power consumption for various RAN splits.
- In **Chapter 5**, we investigate the survivability aspect for C-RAN. In this context, we propose three protection approaches: i) Dedicated Path Protection (DPP) ii) Dedicated BBU Protection (DBP) and iii) Dedicated BBU and Path Protection (DBPP). When designing a resilient C-RAN, we formalize each approach as an ILP problem and solve it over a 5G optical aggregation network.
- **Chapter 6** we continue investigating the survivability aspect for C-RAN. We shift our focus to shared protection schemes showing the saving introduced compared to dedicated protection.
- **Chapter 7** provides conclusions about supplied and open issues with corresponding future work are discussed.

Background on Centralized Radio Access Network 2

This chapter focuses on explaining the concept of C-RAN. The chapter starts with an overview on the mobile network architecture. Then it explores the main differences between C-RAN and D-RAN. Different centralization strategies is presented. The main motivations behind this technique are described, and the critical drawbacks are detailed, mainly related to the restricted requirements of the fronthaul transport network.

2.1 Introduction

The current growth of the mobile traffic as well as the number of connected users drives the academic and industrial research to explore novel mobile network architectures and technologies. Some novel features (e.g., CoMP, HetNets) have been introduced to the currently exist network architectures (e.g., LTE) in order to meet the explosion of the traffic growth. However, the need for radical changes in the current network architecture is considered as the promising solution to face the revolution in the data traffic volumes. C-RAN is introduced as one of those novel 5G paradigms which evolves the mobile network architecture. C-RAN introduces encouraging savings in the network total cost and the energy consumption. Despite the attractive advantages, C-RAN also comes with its own challenges in fronthaul transport network. It is expected that 5G will be commercially deployed in 2020, therefore, a crucial need to understand concept behind C-RAN, how it

will be deployed?, the advantages and disadvantages. In this chapter we try to answer the previous questions. In Section 2, we present a preliminary description of the traditional mobile network architecture. In Section 3, we detail how the base station evolved until we reach the centralization era. In Section 4, we highlight the differences between traditional D-RAN and C-RAN. In Section 5, description of the possible BBU pooling strategies is provided. In Section 6, the main advantages and challenges of C-RAN are then detailed described.

2.2 Mobile Network Architecture

Mobile network is divided into three parts: Radio Access Network, Backhaul Network and Core Network.

2.2.1 Radio Access Network (RAN)

The Radio Access Network (RAN) includes the Base Stations (BS) which perform the radio transmission and digital processing functions of layer 1, 2 and 3. During the last decades the RAN architecture changes from comprising Base Transceiver Station (BTS) and Base Station Controller (BSC) in the 2G to a flat architecture in the 4G which comprises e-NodeB. However, operators tends to deploy different Radio Access Technologies (RATs) -i.e., 2G, 3G and 4G- in the same serving area with an efficient Radio Resource Management (RRM) schemes to take advantage of the available system resources. The main function of the RAN is to connects the end user to the mobile network via the so-called air interface between the mobile device and the BS. The air interface is deployed by the means of multiple access protocols, e.g, Code Division Multiple Access (CDMA) and Orthogonal Frequency Division Multiple Access (OFDMA). Moreover, the RAN performs some higher-layer radio-access functions such as RRM. “Handover” is one of the fundamental function of the RANs which can be explained as follows. Each BS manages UEs belonging to a specific coverage area, denoted as “cell”, and the RAN coordinates the procedures for user mobility, i.e., allowing UEs to move across adjacent cells, without losing data connection.

Base stations are placed into premises denoted as cell sites, whose geographic coordinates are influenced by many different factors, most notably coverage, capacity planning and infrastructural/costs constraints. To save costs, a consolidated practice is implementing more than one BS into a single cell site, thus dividing the coverage area into

up to three cells, denoted also as sectors. For the same reason, BSs of different RATs and different mobile operators often share the same cell site.

2.2.2 Backhaul Network

The Backhaul Network performs traffic aggregation and transport between the RAN and the Core Network. For this reason, its architecture and implementation can be almost agnostic with respect to radio access and core architectures, so they are not covered by RAT standards, given that adequate transport capacity and QoS (e.g., latency) requirements are guaranteed.

2.2.3 Core Network

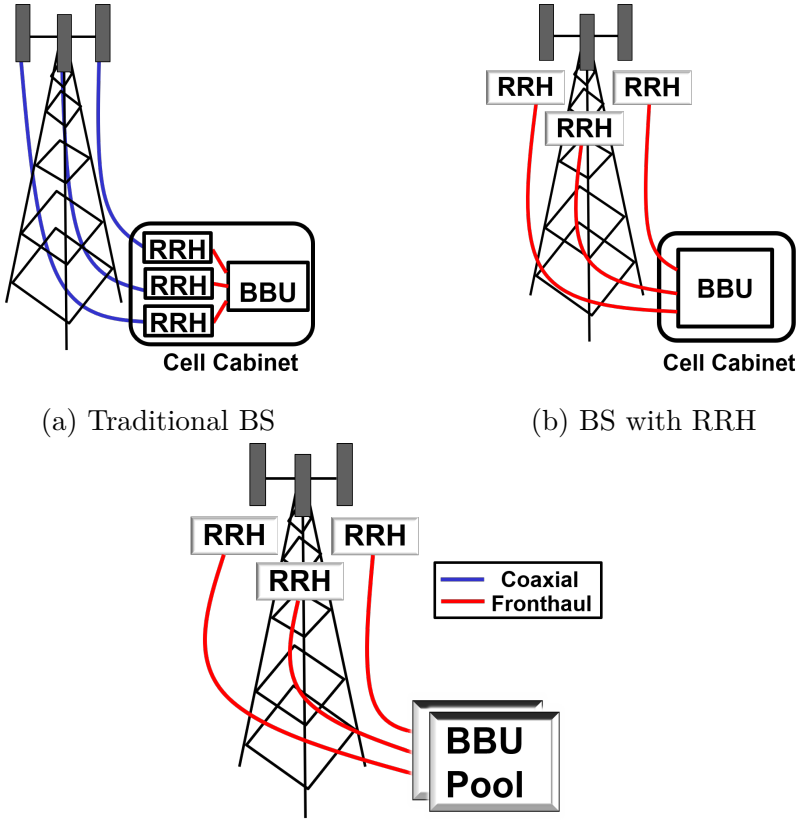
Finally, the Core Network is in charge of all remaining non radio-access related functions and acts as a gateway towards all other mobile and fixed networks, i.e., towards the Internet. Some core network functions and interfaces are standardized too, in most of cases accordingly to the adopted RAT.

2.3 Mobile Network Evolution Towards C-RAN

In the traditional (distributed) RAN the base station comprises two models

(i) **RRH:** provides wireless signal coverage, by transmitting Radio Frequency (RF) signals to UEs in the downlink and forwarding the baseband signals from UEs to the BBU pool for centralized processing in the uplink. RRHs perform RF amplification, up/-down conversion, filtering, Analog-to-Digital/Digital-to-Analog Conversion (ADC/DAC) of the baseband signals. By conducting most signal processing functions in the BBU, RRHs can be relatively simple, and can be distributed in a largescale scenario in a cost-efficient manner.

(ii) **BBU:** performs functions of layer one, two and three (L1, L2, L3). L1 includes functions such as modulation and (de)modulation, sampling, (de)quantization, (de)coding. L2 includes transport and MAC functions, while higher layers include control and network functions



(a) Traditional BS

(b) BS with RRH

(c) Centralized BS

Figure 2.1: Base Station evolution

The way in which the previous mentioned modules (RRH and BBU) are located in the base station has evolved along years. In the following subsections we will overview the main three steps through which the architecture has been evolved.

2.3.1 Traditional Base Station

This architecture is shown in Fig. 2.1a. In this architecture, the BBU and RRH are installed -either in one module or each one on separate module- in the same cell site cabinet. The RRH is connected to the antenna through coaxial cable. In general, this architecture experience high power loss in the coaxial cable depending on the distance between the antenna and the cell cabinet. This type of architecture was employed in 1G and 2G mobile networks.

2.3.2 Base Station with Remote Radio Head (RRH)

This architecture is shown in Fig. 2.1b. In this architecture, the BBU remains in the cell cabinet while the RRH is placed besides the antenna. The main advantage of this solution is that the RRHs can be placed on rooftops to reduce the air conditioning energy consumption. The BBUs can be placed in a more convenient site with lower rental and maintenance costs. The Common Public Radio Interface (CPRI) [13] protocol is used as a radio interface protocol for In phase/ Quadrature (I/Q) data transmission between RRH and BBU. CPRI requires very high data bit rate and very low latency. Each RRH is statistically assigned to one BBU. This architecture is first deployed in 3G networks.

2.3.3 Centralized Base Station

This architecture is shown in Fig. 2.1c. In a C-RAN architecture the BBUs are not only separated from the RRHs, but they are located in a centralized unit, the BBU pool, capable to host several BBUs. This way the housing facility expenses and energy consumption can be considerably reduced. Moreover, a centralized unit provides a common communication channel between the BBUs. This can be exploited to perform coordinated processing. A further step is taken by implementing a virtualized BBU pool consisting of general-purpose processors for baseband processing. General-purpose processors can dynamically be assigned to different RRHs. This allows performing load balancing and efficient resource utilization. The term C-RAN stands at the same time for centralized, clean, cooperative and cloud RAN. In the following sections we will focus on the concept of centralization, how to implement, advantages as well as the disadvantages.

2.4 Centralized Radio Access Network

After mentioning the various ways to locate the RRH and the BBU, in this thesis we focus on the third option mentioned in Section 2.3.3. In this section, we elaborate more on the difference between the distributed RAN considering the BS described in Fig. 2.1b and the centralized RAN considered.

As mentioned previously, in a traditional Distributed RAN (D-RAN), the base station comprises two modules, the Remote Radio Head (RRH) for transmission and reception of radio signals, and the

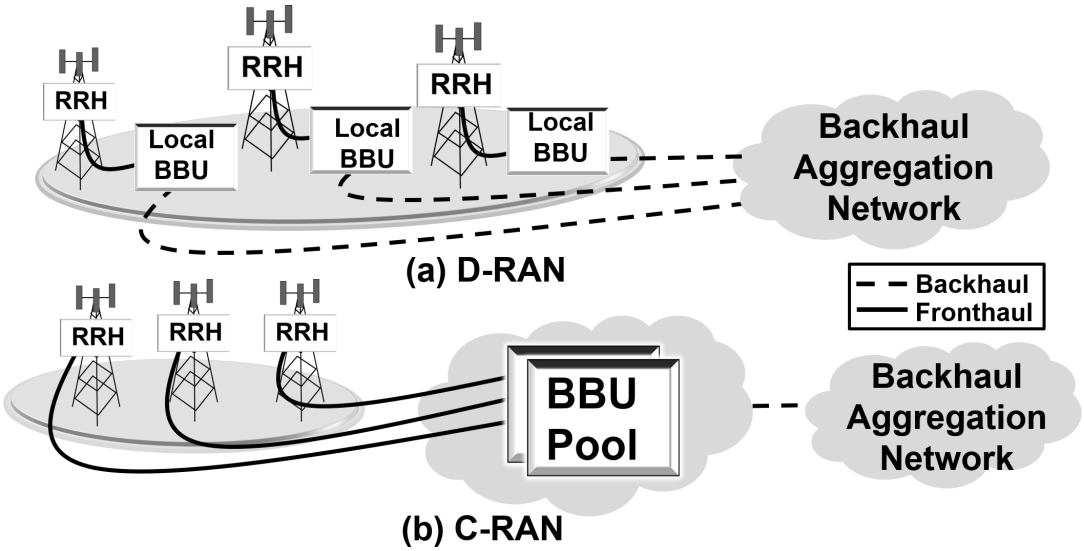


Figure 2.2: Distributed RAN VS Centralized RAN.

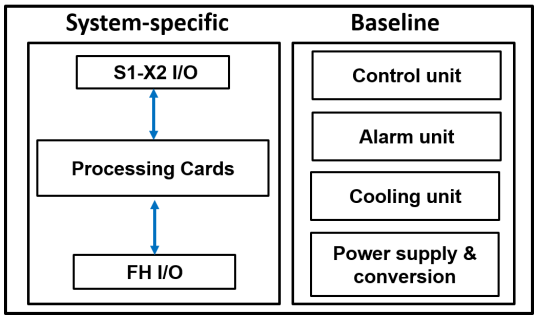


Figure 2.3: Local BBU architecture.

Baseband Unit (BBU) for baseband signal processing; both are located at the same physical site. Every base station has a dedicated housing facility, which is not shared with other base stations as shown in Fig. 2.2a. Subsequently, investment and maintenance costs together with power consumption increase linearly with the number of base stations. Given the rapid growth of traffic demands forecast for the next years, network densification in D-RAN represents a valid solution to address the required network capacity. However, D-RAN does not scale well with the number of base stations, for the reasons aforementioned. New architectures are necessary in order to obtain a scalable solution to limit cost, power consumption and to fulfill the curving ahead traffic demands. C-RAN introduces a valid alternative to D-RAN, since it can provide both scalability and power and cost efficiency. The main

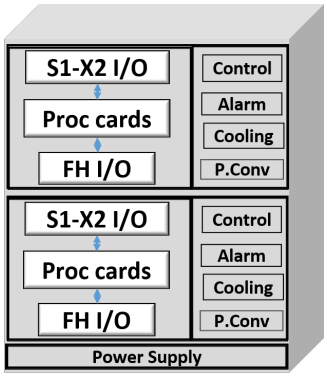


FIGURE 2.4: Stacked BBUs architecture.

idea of C-RAN is to centralize the digital processing units in a single location and to share them among sites in a virtualized BBU pool as shown in Fig. 2.2b. The centralization allows sharing maintenance costs and power consumption among several digital processing units. Moreover, the baseband resources in the pool can be shared among the base stations through a virtualization process.

2.5 BBUs Centralization Strategies

In this section we focus on two different BBUs centralization strategies namely: BBU stacking and BBU pooling. Before presenting the different centralization strategies we start by showing the strcutre of the basic BBU (local BBU). The schematic model of a local BBU architecture (taken from [14]) is shown in Fig. 2.3 . 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 include all the necessary interfaces to exchange data (I/O), such as S1, X2, fronthaul (generally using CPRI) interfaces, and two types of processing units. The first one is the BaseBand (BB) processing unit, which is the computing element implementing Layer 1 (L1) and Layer 2 (L2) functionalities. The second processing unit is general processing card, which implements Layer 3 (L3) functionalities. This basic equipment is usually deployed locally at the base station and for this reason can be referred as the local BBU. The power consumption for such equipment mainly consists in summing up the power consumption of every block.

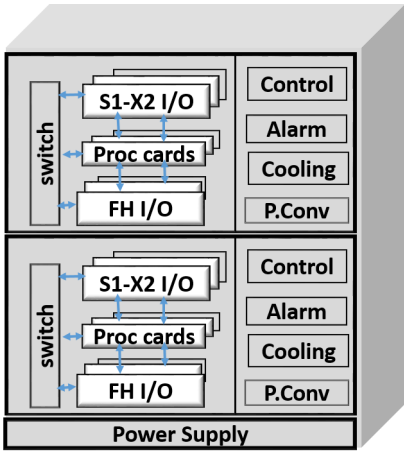


Figure 2.5: BBU pool architecture.

2.5.1 BBU Stacking

The first BBU centralization strategy is BBU stacking, in which original BBUs are simply placed in centralized physical location, rather than cell sites, without changes on their implementation. Thanks to the locating the BBUs in a single physical location, collateral systems, i.e., cabinets, racks, power supply, cooling system, can be re-implemented in a way such that relevant energy and costs savings can be achieved. In this architecture, each BBU is assigned to one cell site, therefore, their number is remained unchanged.

Fig. 2.4 shows a possible implementation of BBU stacking where several BBUs mounted on a rack and the power supply unit shared among all the BBUs in the rack. Note that still it is required to have a power conversion unit in every shelf. The main savings compared to local BBU are in shared power supply unit and the shared OpEx for the whole centralized location. Note that for a specific number of local BBUs, the OpEx obtained is by multiplying this number of BBUs by the cost of one BBU, while the same number of BBUs in one location would require less rental cost because of sharing many BBUs in one rack, depending on how many BBUs could fit in a rack.

2.5.2 BBU Pooling

The second centralization strategy is BBU pooling. As shown in Fig. 2.5, the placement of BBUs at the pool is completely different from their placement at the cell sites. The main feature is that they share some portion of their hardware resources. For this reason, we refer

to the BBUs in a single location as a single BBU "pool". Resource pooling can be of two kinds: static or dynamic. Static pooling occurs when a processing function performed by a single resource element, rather than replicating it across many separate elements. For instance, some low-level physical layer functions requiring intensive vector-based processing (typically joint multi-cell and multi-user algorithms) are often performed by dedicated hardware. By consolidating such hardware, it is possible to improve computational performance and in some cases reduce energy consumption. The pooling is static, because these functions are fixedly assigned to hardware, at the design stage. Dynamic pooling consists in allocating computational resources "on demand" for processing signals of different cells. In this way, to some extent, resources usage can be adapted to the load of each controlled cell, i.e., reserving more resources to high-loaded cells with respect to low-loaded ones. In this way, differently from a traditional BBU exhibiting almost constant energy consumption as function of the load, the BBU pool energy can better scale with it. Adapting the RAN consumption to the traffic load is seen as the most promising way of improving energy efficiency.

2.6 C-RAN Advantages

- **Efficient Network Capacity Utilization:** The capacity enhancement arises from the fact of that C-RAN allows implementing scheduling techniques for interference reduction. The BBUs are provided with a low latency communication channel through which they can jointly contribute on interference reduction. CoMP (Coordinated MultiPoint) techniques for interference reduction have been proposed. Strict synchronization and low latency requirements must be satisfied.

- **Efficient Energy Consumption:** Considerable energy savings can be reached by reducing the number of housing facilities. It is estimated that up to 28% of energy savings compared to traditional D-RAN architecture can be achieved depending on the capacity of the BBU pool in terms of number of cells [15]. Energy consumption from air conditioning and power supplies is reduced because shared among several BBUs in the pool. Moreover, a fewer number of BBUs is needed compared to a traditional D-RAN. The virtualization process in BBU pools allows to se-

lectively turn off unneeded BBUs without compromising a 24/7 service commitment.

- **Efficient Resource Utilization:** In D-RAN the BBUs are dimensioned to accommodate peak load. However, peak traffic occurs for short period during the day, hence inefficient resource utilization occurs. The waste of processing resources can be solved through a virtualized pool solution with load balancing and resource sharing. Load balancing allows overloaded BBUs to migrate the traffic to under-loaded units. Resource sharing allows the overall capacity required in the pool to be smaller than the sum of the single capacities of the base stations due to the enabled multiplexing gain, therefore the number of BBUs can be reduced. In [15], it is shown how the number of BBUs can be reduced by 24% in a metropolitan area with respect to the traditional D-RAN architecture.

- **Easier Network Maintenance and Upgrade Operations:** BBUs centralization in a single location reduces the cost of maintenance required. These operations are therefore considerably facilitated. Moreover, the configuration and maintenance of many servers and routers, in D-RAN deployment, is a complex challenge. Software Design Network (SDN) offers a simplified solution for this complex challenge [16]. Finally, the BBU pool can be implemented to support multiple communication standards and systems, this increasing the network flexibility.

2.7 C-RAN Disadvantages

In this section, there are some main challenges need to be addressed in order to be able to implement C-RAN architecture. Challenges are mainly related to fronthaul traffic requirements, which is introduced to be exchanged between RRHs and BBU pools, whose important features are described in [11] for more extensive discussion.

- **Fronthaul Bandwidth Requirement:** One of the main problems is that the bit rates for the traffic transported on the fronthaul links do not scale with the varying traffic load condition of the cell, resulting in a fully non-elastic traffic. To make a practical example, a typical LTE baseline scenario can be a 20 MHz single sector, 2x2 Multiple Input Multiple Output (MIMO) configuration resulting in 2.5 Gbps bit rate in fronthaul link [17]

regardless the number of users served by this sector. Given that the one BBU pool is connected to more than one cell site, the amount of data carried on such fronthaul links will be very huge.

- **Latency Requirements:** Latency and jitter requirements must be strictly supported, in addition to high bandwidth and cost efficiency. Stringent timing conditions for some physical layer procedures between BS and UEs are specified by RAT standards. Most of them explicitly pose bounds on the latency due to internal processing of radio frames by the BS. In BBU hotelling the BS functions are actually spread between BBUs and RRHs, potentially located very far apart from each other, therefore the "fronthaul latency", i.e., the delay contribution due to the transport of fronthaul signals along the RAN infrastructure, has a relevant impact on the total latency budget inside the BS. Such latency can be further divided into two contributions, firstly, adaptation of fronthaul signals into the RAN transport service and additional functions required by optional lower-layer transport technologies, secondly, signal propagation along the RAN.

2.8 Conclusion

This chapter has introduced the main motivations behind the novel centralized RAN, together with its potential advantages and critical issues for implementation. Promising savings in terms of computational resouuces and energy are expected to be enabled by C-RAN. The main disadvantage comes from the introduction of a new kind of network traffic, called fronthaul, which has radically different features respect to backhaul traffic and enforces strict transport requirements. We present an overview on the mobile network architecture. Moreover, we present the main difference between the C-RAN and D-RAN. How to estimate the computational and energy savings? how and where and where to place the BBUs taking into account the resilience aspect are important questions in the design of C-RAN architectures. Some of such questions will be analyzed and answered throughout the following chapters.

Multiplexing Gain Analysis for 5G RAN Function Splits 3

This chapter focuses on estimating the computational savings enabled by centralizing the BBUs and its influence on the BBUs deployment cost. Moreover, it tackles the trade-off between the centralization gains and the fronthaul requirements through exploring different "functional RAN splits".

3.1 Introduction

Unlike traditional D-RAN architecture, where the processing units are placed in cell sites together with the RF antennas, in C-RANs multiple BBUs are placed in a central pool connected to the RRHs – deployed according to a given radio planning strategy - through an optical transmission network that supports the so called "fronthaul links". The "centralization" of baseband units enables substantial savings of computational resources (what we call "multiplexing gain" in this work) as well as TCO reduction. Despite the advantages provided by C-RAN, huge fronthaul bandwidth is needed with very low latency, so a trade-off arises between the increased transport costs of fronthaul links and the potentially decreased BBU costs associated with centralization. "RAN functional splits" is introduced as a compromised solution that mitigates the fronthaul requirements while enjoying the centralization benefits. In this work, we focus on the sharing of processing (BBU) facilities in C-RAN, as processing centralization promises to enable a significant multiplexing gain. In addition, we show the cost reduction

for BBUs deployment in C-RAN compared to D-RAN. Finally, We present different RAN splits showing the trade-off between relaxing the fronthaul requirements and the reduction in the centralization gains.

3.2 Related Work

Multiplexing gain that arises from the aggregation of multiple BBUs in a central pool has also been investigated. Ref. [18] presents a multi-dimensional Markov model to derive the multiplexing gain. Ref. [19] estimates the multiplexing gain based on a simulation study in the case of multiple sectors aggregated into a single cloud BS. Ref. [20] presents traffic simulation experiments to evaluate the multiplexing gain in WiMAX BSs under different traffic conditions. In [21], 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 [22] shows that reduction in the computation resources is up to 70%. Ref. [23] 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.

Many studies have been performed to investigate the cost reduction achieved by C-RANs from the perspective of Capital Expenditure (CapEx) and Operational Expenditure (OpEx). Authors in [2] propose a network architecture exploiting Software Defined Network (SDN), Network Function Virtualization (NFV) and C-RAN technologies to show their influence on CapEX, OpEx and Total Cost of Ownership (TCO). They estimate a significant 68 % reduction compared to the traditional D-RANs. However, the authors do not provide a detailed model accounting for baseband pools, and consider their cost as a fixed value, neglecting the influence of eNBs aggregation. Ref. [3] proposes a cost model for the central baseband processing pool in C-RAN based on different strategies to construct it, namely: stacking, pooling and cloud-RAN, but also in this case, a fixed multiplexing gain is assumed to evaluate the reduction in the computational resources.

Several works have been conducted on functional splits. In [24] 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 [25], [26] and [27]. In [28], the authors present a model to calculate the fronthaul bandwidth and the

Table 3.1: 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

computational resources required for different functional RAN splits. In [29], authors present a comparison of the achievable pooling and CoMP gains for different RAN splits. Ref.[30] quantifies experimentally the transport-network capacity requirements of three different RAN splits, in terms of control plane and user plane data. Ref. [31] 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. [32] 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.

3.3 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 [33], [34][1], i.e.:

$$P_{R_{uk}} = P_{T_k} - (128.1 + 37.6 \log(d_{uk})) \tag{3.1}$$

  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 [35] 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 [35]:

$$r_{us} = \frac{c_n \log_2 m_n}{T_{SB}} N_{SB} \tag{3.2}$$

[1]Note that the pathloss equation is given by: $L(dB) = 40(1-4\times10^{-3})log(d_{uk}) - 18log(Dhb) + 21log(f) + 80dB$ 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 $L(dB) = 128.1 + 37.6log(d_{uk})$.

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 threshold ($SNR_{Threshold}$) as stated in [36]. 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 [37]. Then, the sets of allocated SBs and used MCSs for user u are obtained by maximizing the following quantity

$$\sum_{u=1}^{|U|}\sum_{s=1}^{|S|}\sum_{n=1}^{M_u} w_{usn}\frac{c_n \log_2 m_n}{T_{SB}}N_{SB} \tag{3.3}$$

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 Eq. (3.3), 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 [35]. Maximization of Eq. (3.3) is obtained as presented in [35]. MIMO mode and spatial streams are obtained knowing MCSs index [38]. Hence, in a given TTI, we know the SBs allocated to user u, used MCSs, MIMO mode and spatial streams.

- **Computational Effort:** The computational effort for user u at TTI t can be defined as [19]:

$$CE(u,t) = \left(3A_{u,t} + A_{u,t}^2 + \frac{1}{3}M_{u,t}C_{u,t}L_{u,t}\right)\frac{R_{u,t}}{10} \tag{3.4}$$

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

$$CE_{DRAN_{total}} = \sum_{u}\sum_{t} CE_{DRAN}(u,t) \tag{3.5}$$

$$CE_{CRAN_{total}} = \sum_{u} \sum_{t} CE_{CRAN}(u,t) \tag{3.6}$$

where $CE_{CRAN}(u,t)$ is computational effort computed by Eq. (3.4) for a user served by centralized RAN and $CE_{DRAN}(u,t)$ is computational effort computed by Eq. (3.4) for a user served by distributed RAN.

- **Multiplexing Gain Calculations:** 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. (3.5)) and C-RAN (as in Eq. (3.6)) 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

$$\beta = \frac{CE_{DRAN_{total}} - CE_{CRAN_{total}}}{CE_{DRAN_{total}}} \tag{3.7}$$

3.4 Cost Model

After calculating the multiplexing gain introduced by the centralization of BBUs, we move forward and present an analytical cost model for the deployment of centralized an distributed BBUs. This model shows how the obtained multiplexing gain affects/reduces the cost. First, we introduce the cost model for baseband processing in a D-RAN. To show the basic building blocks of a "local BBU", i.e., a BBU deployed at the cell site, we refer to Fig. 2.3. The units for control, alarms, cooling and power supply are considered as baseline blocks. Interfaces for S1-X2, fronthaul (e.g. CPRI) and data exchange (I/O) as well processing units are considered system-specific units. The total cost

C_T for such building blocks is calculated as follows,

$$C_T = C_{BL} + C_{PS} + C_{FH} + C_{S1-X2} + C_{PC} \tag{3.8}$$

where C_{S1-X2} is the cost of a S1-X2 interface, C_{FH} is the cost of the fronthaul interface, and C_{PC} is the cost of the processing cards, C_{BL} is cost of the baseline unit and C_{PS} is the cost of the power supply including power conversion. In BBU pooling, the main feature is that multiple BBUs share the computational. We refer to Fig. 2.5 to show the architecture of a BBU pool. The total cost for BBU pool is calculated as follows,

$$C_T = s\,(C_{BL} + C_{LS}) + s\,\nu\,(C_{FH} + (1-\beta)C_{PC}) + s\,\gamma\,C_{X1-S2} + r\,(C_{PS} + C_r) \tag{3.9}$$

where, C_r is cost of the empty rack, s is number of shelves, r is number of racks, ν number of aggregated sites per shelf, β is multiplexing gain (β<1) and γ is number of S1-X2 interface per shelf. Moreover, with the pooling strategy, the number of fronthaul interfaces per shelf is equal to the number of cells aggregated per shelf. The number of processing cards per shelf equals the number of fronthaul interfaces per shelf multiplied by the multiplexing gain. While the ratio between the number of S1-X2 interfaces to the number of processing cards is 1:6. 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)$\rceil$.

3.5 RAN Functional Splits

Despite the computational and cost savings (explained in the previous sections) arises from the deployment of C-RAN, restricted fronthaul requirements in terms of very high bit rates and very low latency still impede the implementation of C-RAN. Therefore, mobile operators are seeking new solutions to relax the fronthaul constraints through a more flexible distribution of baseband functionalities between RUs and DUs [39]. 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. This splitting will be clarified in detail in this section.

3.5.1 3GPP protocol Stack

In order to introduce the different RAN functional splits, we overview the various functions, grouped according to the protocol layer, as defined in 3GPP LTE RAN[40].

(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.
- **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.

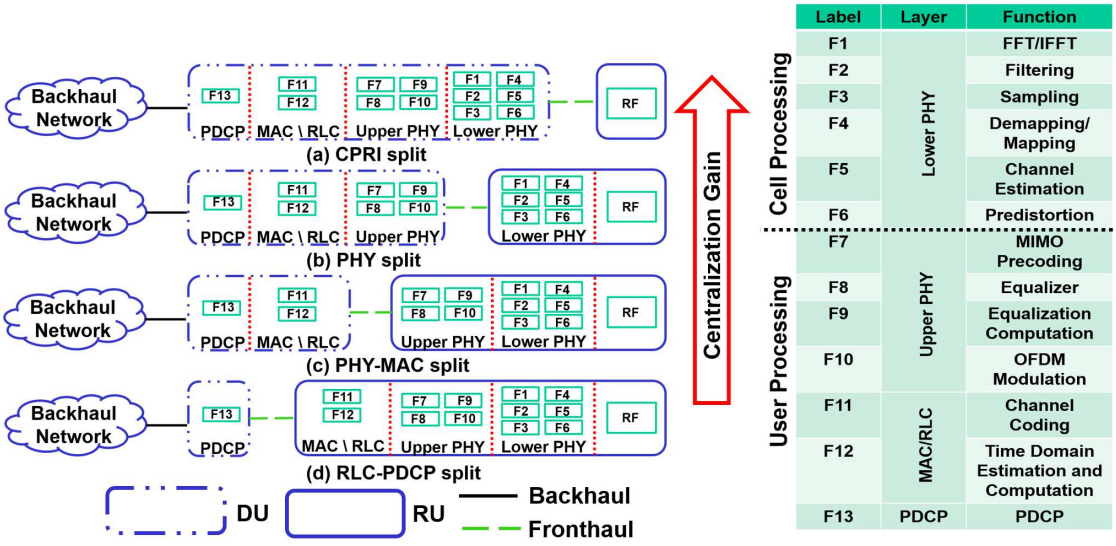


FIGURE 3.1: RAN split options and corresponding mapping of network functions.

3.5.2 Split Options

As mentioned previously, the separation of the functions can be applied on any protocol layer, or on the interfaces between layers. Therefore, we present four possible functional RAN splits depicted in Fig. 3.1 as follows

CPRI split: Fig. 3.1a shows the CPRI split (also referred as full centralization) , 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[26]) for physical layer processing.

PHY split: Fig. 3.1b 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 is relaxed to 2ms [26]. 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. 3.1c. 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 [26].

RLC-PDCP split: RLC-PDCP split is shown in Fig. 3.1d. 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) [26] since all the scheduling functions (e.g., MIMO precoding and OFDM modulation) are distributed at the cell site.

3.5.3 Computational Effort and Multiplexing Gains the RAN Functional splits

The calculation of computational effort for functional RAN splits: PHY split, PHY-MAC and RLC-PDCP, differs from the CPRI split option (calculated in Section 3.3), since 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. (3.4), 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

$$\sigma = \frac{GOPS_{DU}}{GOPS_T} \tag{3.10}$$

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. (3.10) 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,

Table 3.2: Complexity Factor Values

RAN Functional Split	**Complexity Factor** (σ)
PHY	0.6
PHY-MAC	0.3
RLC-PDCP	0.1

the computational effort CE_{rs} for user u at TTI t in the RAN split rs can be defined as

$$CE_{rs}(u,t) = \sigma CE(u,t) \tag{3.11}$$

where, σ is summarized in Table 3.2

And, the multiplexing gain for the different RAN splits can be calculated as follows again as in Eq. (3.7) but considering the computational effort as in Eq. (3.11) as follows

$$\beta_{rs} = \frac{CE_{DRAN_{total}} - CE_{rs}}{CE_{DRAN_{total}}} \tag{3.12}$$

3.6 Illustrative Numerical Results

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

3.6.1 Evaluation Settings

We consider three different geographical type areas (geotypes), namely, Dense urban (D), Urban (U), and Sub-urban (S). The number of cell

Table 3.3: Features of the Considered Geotype.

	Dense Urban	**Urban**	**Suburban**
Number of Sites per Km²	4	1.5	0.2
Total Area to Accommodate 12 site [km²]	3	8	60
Number of User per km²	3000	1000	500
Total Number of Users	9000	8000	30000

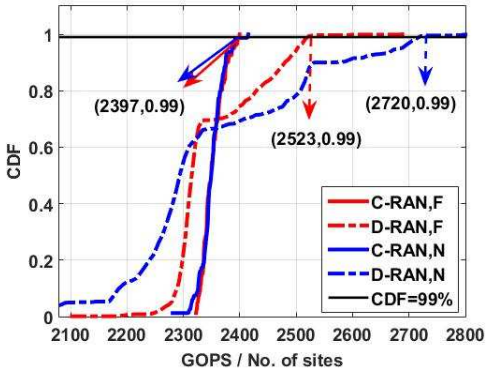


Figure 3.2: Cumulative distribution function for computational effort in dense urban area considering normal and uniform users' distributions.

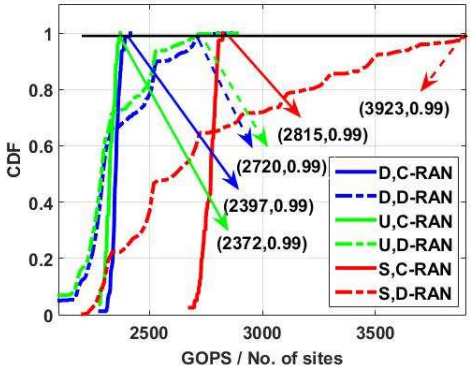


Figure 3.3: Cumulative distribution function for computational effort considering three different geotypes with normal users' distribution.

sites per unit area and number of users per unit area are taken from [41] and [42], respectively, and reported in Table 3.3.

3.6.2 C-RAN Multiplexing Gain Assessment (CPRI)

We start observing the statistical properties of the computational effort calculated using Eq. (3.4) to calculate the multiplexing gain and show the impact of the users distribution, different geotypes and the different pool dimension, for different functional RAN splits.

Fig. 3.2 shows the cumulative distribution function (CDF) of the computational effort (CE) in a dense-urban area, where users are

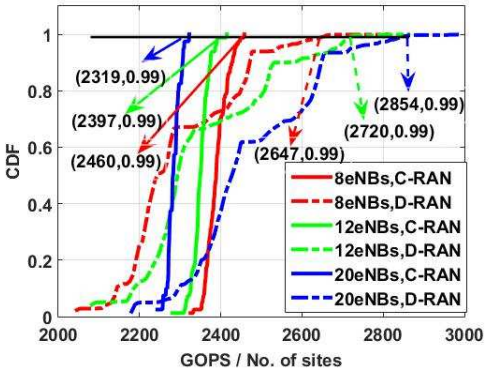


FIGURE 3.4: Cumulative distribution function for computational effort in dense urban area considering different pool dimensions with normal users' distribution.

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

$$\beta = \frac{2720 - 2397}{2720} \times 100 = 11.8\% \tag{3.13}$$

while multiplexing gain for uniform distribution is calculated as

$$\beta = \frac{2523 - 2397}{2523} \times 100 = 5\% \tag{3.14}$$

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. 3.3 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 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[2] 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. 3.4 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.

3.6.3 C-RAN Cost Assessment

With BBU pooling, multiple processing cards in the same shelf are interconnected by a low-latency switch and shared among multiple eNBs. In spite of added hardware complexity compared to basic BBU, lower cost is expected due to statistical multiplexing gain. Consider a shelf size which supports 12 aggregated eNBs ($r = 1, \mu = 12$) and, a rack with 42 Rack Units (RU) which accommodates the aggregated eNBs ($r = 1$). The cost assumptions in [14] are considered, where the costs are normalized to a cost of power consumption per MWh. Fig. 3.5 shows total cost of baseband pool for different geotype areas

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

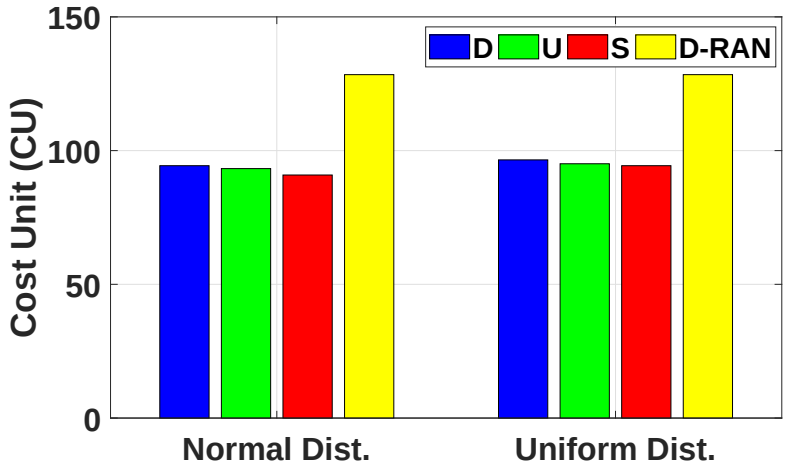


FIGURE 3.5: Cost of baseband pool.

and compares it with the basic BBU cost. As it is mentioned earlier, the statistical multiplexing gain in our model is maximized in the sub-urban area and minimized in the dense urban area. Therefore, as expected, baseband pool in sub-urban area has the lowest cost. For normal users distribution, pooling savings up to 26% can be achieved in dense urban area compared to the basic BBU, while in sub-urban area 30% pooling savings are obtained. For uniform users distribution, pooling saves of 24% can be achieved in dense urban area compared to the basic BBU, while in sub-urban area 27% pooling saves can be achieved. Note that the cost of the fronthaul network is expected to be much higher in a suburban area than in a dense-urban area due to the much larger distances to be covered by fibers.

3.6.4 RAN Functional Splits Gain Assessment

Finally, in order to capture the multiplexing gains obtained for different RAN functional splits in the different geotypes, we apply Eq. (3.12) considering 12 eNBs with normally distributed users. The results shown in Fig. 3.6 quantify the decrease of the multiplexing gain due to the adoption of less aggressive functional splits.

3.7 Conclusion

In the light of savings enabled by C-RAN, we propose a multiplexing gain model to capture the processing savings arising from consolidation

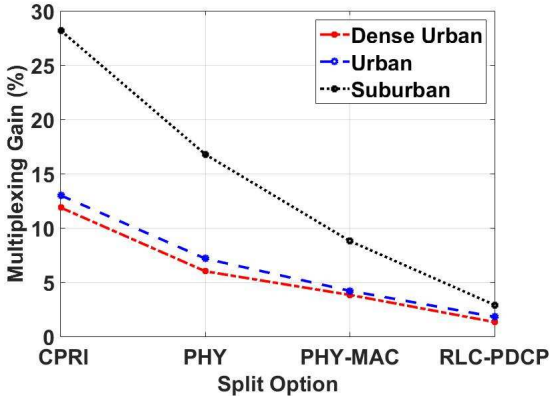


Figure 3.6: Multiplexing gain for different RAN splits considering different geotypes with normal users' distribution.

of compute resources, considering four different functional RAN splits: CRPI, PHY, MAC-PHY and RLC-PDCP. We also propose a cost model to evaluate the cost reduction enabled by C-RAN.

Results prove that the highest multiplexing gains 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.

Power Consumption Analysis for 5G RAN Functional Splits 4

In this chapter we continue exploring the gains enabled by C-RAN. This chapter focuses on power saving arises due to the centralization of BBUs. In order to do that, we present a power consumption model for the C-RAN with the different splits as well as D-RAN.

4.1 Introduction

According to Ericsson's Mobility Report [2] by 2020 over 90 % of the population over six years old will have a mobile phone and smartphone subscriptions are expected to top 6.1 billion, resulting in much higher traffic demands for wireless communications. This growth is obviously accompanied by an increased power consumption of mobile networks. This dramatic increase of the power consumption is a significantly contributing cause towards the creation of carbon dioxide that leads to Green House Gas (GHG) problems, as well as other global warming pollutants. On top of that, it significantly increases OpEx of mobile networks.

For the above mentioned reasons, the power efficiency of the wireless mobile networks is considered as a crucial element that maintains sustainability of future mobile networks. Therefore, the reduction of the mobile network power consumption has become more critical.

Researches on power consumption of mobile networks (mobile equipment, base stations, and core network) reveal that 80 % of the power needed in mobile network operation is consumed at base stations.

In order to keep the power consumption evolution of mobile networks under control, the base stations are thus the main center of attention for optimization.

This chapter focuses on estimating the power savings enabled by C-RAN compared to D-RAN. In Section 4.2, we discuss the related work regarding the energy savings in 5G. In Section 4.3, we present an overview on power consumption contributors of the base station. In Section 4.4, we introduce our power consumption model for D-RAN. While in Section 4.5, we introduce our power consumption model for C-RAN. In Section 4.6, we show the illustrative numerical results. Finally, in Section 4.7 we conclude the work.

4.2 Related work

Energy consumption of mobile networks has been subject of intensive research. Ref. [43] 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. [44] identifies a significant power saving by introducing two different downlink transmission strategies (namely, data-sharing and compression strategy) in C-RAN. Ref.[45] models the power savings achieved by applying dynamic BBU-RRH mapping, showing that, compared to D-RAN, 70% power savings can be achieved. Ref.[46] evaluates the power savings achieved in C-RAN by applying a cooperative transmission scheme with low computational complexity to mitigate interference. In [47], 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.

4.3 Base Station Power Components

In order to capture the power consumption of a centralized base station, it is necessary to be familiar with the architecture of base station. This section presents the general power consumption contributors for a base station. The main power contributors [48] are listed as follow.

- **Power amplifier** plays a vital role in power modeling. It amplifies the electrical signal from DRX (driver receiver) for trans-

mission through antenna interface. It is hard to capture the power amplification measurements by some specific numbers, so it is captured by presenting output power measurements versus consumed power.

- **Analog front-end** comprises amplifiers and filters (analog baseband and RF), up/down-conversion mixers, frequency synthesizer, and digital to analog/analog to digital converter (DAC/ADC).
- **Digital signal processing** comprises digital predistortion, digital compensation of system non-idealities, baseband filtering, up-/down sampling, FFT/IFFT, modulation/demodulation, channel encoding/decoding, channel estimation, synchronization, processing for MIMO (MIMO coding), and equalization. Base band processing functionalities will be explained later in details.
- **Digital control** comprises platform control processor, protocols for radio link management and configuration, backbone serial link interface, and admission control.
- **Antenna interface** comprises Tx/Rx filters, feeder loss (connection between antenna and transceiver), and antennas. Power supply feeds AC/DC convertor, DC/DC convertors, and active cooling.

4.4 Local BBU Power Consumption

As mentioned previously, the evolution of the mobile network goes towards centralized RAN architecture that introduces a significant power saving, which comes from the centralization of BBUs. Conversely, analogue front-end and power amplification remain distributed for both centralized and decentralized RAN architectures. Therefore, analogue front-end and power amplification have no contributions in power savings since they consume the same amount of power in both architectures. Accordingly, the power consumption of the analogue front-end and the power amplification will be excluded from our power calculations.

This chapter focuses on the power consumption savings enabled by C-RAN compared to D-RAN. For simplicity, we refer to Fig. 2.3 to evaluate the power consumption by dividing the BBU components into 5 parts namely:

(i) Processing Cards

(ii) Basline unit

(iii) Fronthaul interface

(iv) Bachhaul interface

(v) Power supply

The power consumption of a local BBU (shown in Fig. 2.3) can be calculated as

$$P_{BB} = (P_{BL} + P_{FH} + P_{S1-X2} + P_{PC})(1 + \alpha_{PS}) \tag{4.1}$$

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 shown in the following subsections.

4.4.1 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 [49] 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. Table 4.1 shows the complexity values of the baseband functions in GOPS. 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.:

$$P_i = \frac{C_i}{T} \tag{4.2}$$

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 [50]. Note that, the complexity values of any baseband function (mentioned in Table 4.1) are referred to a specific scenario (reference scenario), where,

- Bandwidth is 20 MHz

Table 4.1: Reference Complexity of Digital Components.

Digital baseband function	Downlink	Uplink
Predistortion	10.7	0
Filtering	6.7	6.7
Up/Down sampling	2	2
TD non-ideal estimation/ computation	1.3	6.7
FFT/IFFT	4	4
MIMO processing	1.3	0
Synchronization	0	2
Channel estimation and interpretation	0	3.3
Equalizer computation	0	3.3
Equalization	0	2
Modulation/Demodulation (OFDM)	1.3	2.7
Resource Demapping/ Mapping	1.3	2.7
Channel coding	1.3	8
Control	2.7	1
Network	8	5.3

- System is fully loaded (no hardware deactivation)
- Antenna configuration is single-input-single-output
- Spectral efficiency of 6 bps/Hz is achieved
- Modulation scheme is 64-QAM (Quadrature Amplitude Modulation) and coding rate of 1

For any other scenarios rather than the reference scenario, P_i is scaled using factor Γ_i ($\Gamma_i = 1$ for reference scenario).

$$P_i = \frac{C_i}{T}\Gamma_i \tag{4.3}$$

This factor is determined according to the following parameters:

(i) Bandwidth

(ii) Spectral efficiency

(iii) Number of antennas

Table 4.2: Scaling Exponent For the Digital Baseband Functions.

Digital baseband function	Bandwidth	Spectral efficiency	Antenna	Load	Streams	Quantization
Predistortion	1	0	1	0	0	1.2
Filtering	1	0	1	0	0	1.2
Up/Down sampling	1	0	1	0	0	1.2
TD non-ideal estimation/ computation	1	0	1	0	0	1.2
FFT/IFFT	1.2	0	1	0	0	1.2
MIMO processing	1	0	1	1	1	1.2
Synchronization	0	0	1	0	0	1.2
Channel estimation and interpretation	1	0	1	0.5	0	1.2
Equalizer computation	1	0	3	1	0	1.2
Equalization	1	0	2	1	0	1.2
Modulation/Demodulation (OFDM)	1	0	1	0.5	0	1.2
Resource Demapping/ Mapping	1	1.5	0	1	1	1.2
Channel coding	1	1	0	1	1	1.2
Control	0	0	0.5	0	0.2	1.2
Network	1	1	0	1	0	1.2

(iv) System load in terms of hardware activation and deactivation (sleep-states)

(v) Number of spatial streams that are relevant to the number of antennas

(vi) The quantization (e.g., 4 bits, 16 bits and 24 bits)

To model the impact of each of these six parameters, we introduce scaling coefficients, defined as exponential factors. Scaling exponents influence the power consumption of each baseband processing function according to how important the parameter is for the power scaling. If the scaling exponents is zero, then the power consumption remains constant meaning that this parameter does not affect the power scaling. Table 4.2 provides the mentioned scaling exponents. Γ_i is calculated as follows

$$\Gamma_i = \left(\frac{x_{act}}{x_{ref}}\right)^{S_{i,x}} \tag{4.4}$$

where, each parameter x has reference value x_{ref} and an actual value x_{act}, $S_{i,x}$ is the scaling exponent for the baseband processing function i with respect to parameter x

In order to familiarize the reader with power calculations, let's discuss the following two examples:

- **Example 1:** Assume we want to calculate the power consumption of the baseband processing function "Modulation", given that:
  - Bandwidth = 20 MHz
  - No hardware deactivation
  - Antenna configuration is single-input-single-output
  - Spectral efficiency = 6 bps/Hz
  - Modulation scheme= 64-QAM and coding rate = 1
  - Technology factor = 8/3 [49]

  Then, from Table 4.1 the complexity of the "Modulation" function is 1.3 GOPS. The given parameters indicate that this scenario is a reference one ($\Gamma_{Mod} = 1$). Therefore, the power consumption of the function "Modulation" P_{Mod} is calculated as follow

$$P_{OFDM} = \frac{1.3}{8/3} = 0.4875W \tag{4.5}$$

- **Example 2:** Assume the same example as the previous one but the antenna configuration is 2x2 MIMO (This is not a reference scenario). Therefore Γ_{Mod} does not equal to 1. Changing the antenna configuration changes the number of the antenna and streams. Therefore, to calculate Γ_{Mod} we will use the scaling vector with respect baseband function "Modulation" [1, 0, 1, 0.5, 0, and 1.2] as depicted in Table 4.2. Γ_{Mod} is calculated as follow

$$\Gamma_{OFDM} = \left(\frac{BW_{act}}{BW_{ref}}\right)^{S_{Mod,BW}} \left(\frac{S.E_{act}}{S.E_{ref}}\right)^{S_{Mod,S.E}} \left(\frac{A_{act}}{A_{ref}}\right)^{S_{Mod,A}} \left(\frac{L_{act}}{L_{ref}}\right)^{S_{Mod,L}} \left(\frac{S_{act}}{S_{ref}}\right)^{S_{Mod,S}} \left(\frac{Q_{act}}{Q_{ref}}\right)^{S_{Mod,Q}} \tag{4.6}$$

$$\Gamma_{OFDM} = \left(\frac{20}{20}\right)^{1} \left(\frac{6}{6}\right)^{0} \left(\frac{2}{1}\right)^{1} \left(\frac{100}{100}\right)^{0.5} \left(\frac{2}{1}\right)^{0} \left(\frac{24}{24}\right)^{1.2} = 2 \tag{4.7}$$

Then, the power consumption of the baseband function "Modulation" is calculated as follow

$$P_i = \frac{1.3}{8/3} \times 2 = 0.975W \tag{4.8}$$

Finally, after knowing how to calculate the power consumption of every baseband function for any scenario, the power consumption of the processing card is obtained through the summation of power consumed by each implemented baseband function as:

$$P_{PC} = \sum_i P_i \Gamma_i \tag{4.9}$$

4.4.2 Baseline Unit

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) [50]. 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:

$$P_{BL} = P_{PC}\left(1+\eta_{cool}\right)\left(1+\eta_{dc/dc}\right)\left(1+\eta_{ac/dc}\right) \tag{4.10}$$

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.

4.4.3 Fronthaul

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[1] [41].

4.4.4 Backhaul

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 [49].

4.4.5 Power Supply

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 $\alpha_{PS} = 5\%$ in Local BBUs and $\alpha_{PS} = 10$ % in different RAN splits [14].

4.5 DUs Power Consumption

As we calculated the power consumption of the local BBUs, we move forward in this section and calculate the power consumption of the digital centralized units. The total power consumption P_T of a centralized digital unit (as depicted in Fig. 2.5) is calculated as:

$$P_T = \left(s\left(P_{BL}+P_{LS}\right)+s\nu\left(P_{FH}+\left(1-\beta\right)P_{PC}\right)+s\gamma P_{X1-S2}\right)\left(1+\alpha_{PS}\right) \tag{4.11}$$

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

[1] Although the required transport capacity depends on the considered RAN split, for simplicity we assume constant power consumption in the different functional RAN splits.

gain ($\beta < 1$) and γ is number of S1-X2 interfaces per shelf. Note that, the multiplexing gain β is calculated for different geotypes and RAN splits in the previous chapter.

Moreover, the number of fronthaul interfaces per shelf is equal to ν the number of cells aggregated per shelf. The number of processing cards per shelf equals the number of fronthaul interfaces per shelf multiplied by the multiplexing gain ($\nu(1-\beta)$). 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., $\gamma = \nu(1-\beta)/6$. 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)$\rceil$.

4.6 Illustrative Numerical Results

In this section we compares the DUs power consumption estimated using the power model in Section 4.5, for different functional RAN splits. We show the power savings in each split option compared to D-RAN.

4.6.1 Simulation Settings

In this study, 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 [41] and [42], respectively, and reported in Table 3.3. Moreover, we assume a particular air interface 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). Note that, we refer to CPRI split, PHY , PHY-MAC split and RLC-PDCP split as S1, S2, S3 and S4, respectively.

4.6.2 Power Consumption Assessment

Fig. 4.1 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

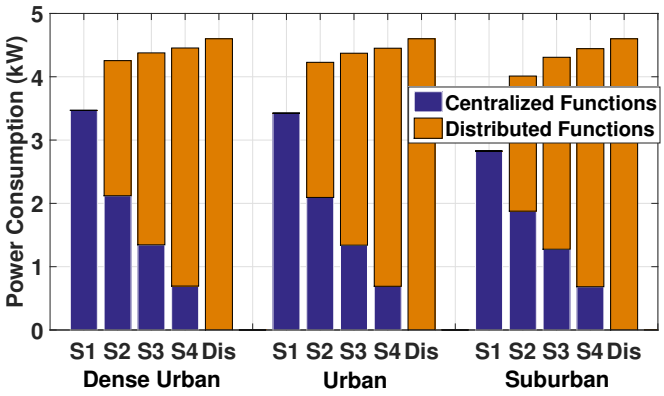


FIGURE 4.1: Power consumption for 12 aggregated cell sites considering different RAN splits in three different geotypes with normal users' distribution.

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. 4.2 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.

4.7 Conclusion

In this chapter, we propose a power consumption model to quantify the savings associated with different RAN functional splits (i.e., CPRI, PHY, PHY-MAC, RLC-PDCP) compared to the traditional one (D-RAN). This model considered different geographical area types (Denseurban, Urban, Suburban), statistical multiplexing gain for each RAN functional split, and particular air interface scenario in terms of: spectrum efficiency, modulation schemes, antenna configuration and spatial streams and system load. The results shows that, power con-

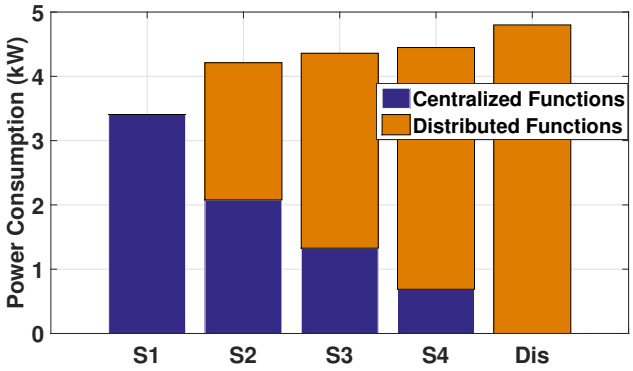


Figure 4.2: Power consumption for 20 aggregated cell sites considering different RAN splits in dense urban area with normal users' distribution.

sumption is inversely proportional to the multiplexing gain. Therefore, the highest power savings is obtained in the sub-urban scenario. On the contrary, in dense-urban environments we obtain the lowest power saving for the same number of cell sites considered.

Resilient BBU Placement in 5G C-RAN over Optical Aggregation Networks 5

In this chapter, we tackle the BBU placement problem in C-RAN taking into account the survivability aspect. We introduce an optimization approach for placing the BBUs in a WDM aggregation network that can deal with network and link failures.

5.1 Introduction

In the previous chapters, we investigate the C-RAN showing the promising savings that arises from centralization. In this chapter, we shift our focus to investigate the challenge of dealing with the network and link failures in C-RAN. Today's businesses increasingly rely on mobile networks, which brings both great opportunities and challenges. One of the critical challenges is resiliency: disruptions due to failures may entail significant revenue losses. Therefore, designing a survivable C-RAN becomes crucial as BBU pool and link failures might cause service outage for large number of users.

In this chapter, we investigate the survivable BBU placement problem and traffic routing for 5G C-RAN deployment. We propose three protection approaches: i) Dedicated Path Protection (DPP) ii) Dedicated BBU Protection (DBP) and iii) Dedicated BBU and Path Protection (DBPP). We formalize each approaches as an ILP problem and solve it over a 5G optical aggregation network.

The remainder of the chapter is organized as follows. Section 5.3 discusses the considerd C-RAN aspects (the network architecture

and the BBU pool computational processing). Section 5.4 discusses the considered protection approaches. Section 5.5 shows the ILP formulations used to solve the resilient problems in the three protection scenarios. Section 5.6 shows the illustrative numerical results. Section 5.7 concludes the work

5.2 Related Work

Despite the C-RAN advantages discussed in the previous chapters, the fronthaul network must be able to support very high bandwidth traffic with very low latency, leading to high transport-network cost. Optical aggregation networks based on wavelength division multiplexing (WDM) are considered a relevant candidate solution to meet the fronthaul requirements [51]. In this context, an optimization placement of BBU pools in the aggregation network is crucial. The placement problem aims to choose a BBU pool location, which meets the latency and bandwidth requirements while maximizing the aforementioned centralization benefits. Many previous works have addressed the placement problem. The authors in [52] formulate an ILP model for the BBU pool placement problem to minimize the power of the aggregation network, while in [51] they investigate the amount of BBU consolidation achieved when using two different transport networks solutions (OTN and overlay) and jointly optimizing the BBUs and the electronic switches placement. Ref. [53] introduces a BBU placement problem for dense small cells over wireless fronthaul network and proposes a heuristic placement algorithm named SWAN to solve the problem. Ref. [54] formulates a Mixed Integer Linear Programming (MILP) for Digital Unit (DU) pool placement optimization problem. The objective of the optimization problem is to minimize the total network cost. Similarly, ref. [32] proposes genetic algorithm to reduce the fronthauling cost through properly splitting and placing the baseband processing functions in the network.

An important aspect in the placement problem is how to deal with BBU pool and link failures, as these failures might cause service outage for a large area with a significant number of users. This has recently motivated novel research on the design of survivable C-RAN. Though survivability in the context of C-RAN and 5G has been rarely discussed in the literature, resilience aspects in the more general optical networks context has been widely investigated. Ref. [55] proposes a multipath protection scheme for data center services in an elastic optical networks based on the importance level of the services. Ref.

[56] proposes a N:1 protection mechanism for Optical Line Terminals (OLTs) aims at minimizing the number of backup OLTs required in a Passive Optical Network (PON). Ref. [57] proposes a cost effective algorithm to minimize the cost of resilient flexible bandwidth optical networks. Ref. [58] presents two models for dedicated and shared path protection against a single link failure in elastic optical networks. Ref. [59] introduces an efficient restoration mechanisms to ensure service resilience in 5G cloud-based mobile network.

More related to our work, ref. [60] defines a protection problem for cloud radio access network against BBUs and link failures. The authors present different approaches based on 1+1 dedicated path protection and 1+1 virtual machine replication through an ILP algorithm, but the authors have not considered delay and link capacity constraints. Ref. [61] proposes a heuristic algorithm to connect each RRH to two BBU pools, primary and backup, while reducing the number of the backup BBU ports among the RRHs. Then in [62], the authors extend the work and formulate the problem as an ILP algorithm and compare it to the aforementioned heuristic algorithm. Ref. [63] formulates an ILP problem named cost-resilience BBU selection, where a mobile network operator has to select BBU equipment from several cloud providers with different failure probability and cost. The objective function minimizes the BBU pool processing power and maximizes the resiliency. So far, to the best of our knowledge no study has considered the latency and link capacity constraints for the fronthaul network in survivable BBU placement problem.

5.3 Centralized radio access network

As already mentioned in Section 6.1, C-RAN introduces a significant computational, cost and power savings which come from the centralization of BBUs. Although the aforementioned advantages, C-RAN requires a high-capacity and low-latency access/aggregation network to support fronthaul traffic. In the following, we will discuss the considered network architecture and the computational processing required by a BBU pool in terms of GOPS.

5.3.1 Network architecture

In this paper, we assume Opaque network architecture as in [52], where the BBU pool can be placed at any node. For intermediate nodes, this implies they become active and equipped with an electronic switch that

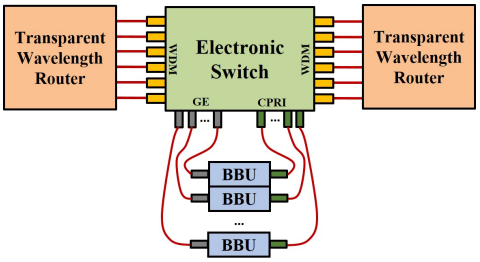


Figure 5.1: BBU pool node architecture [51]

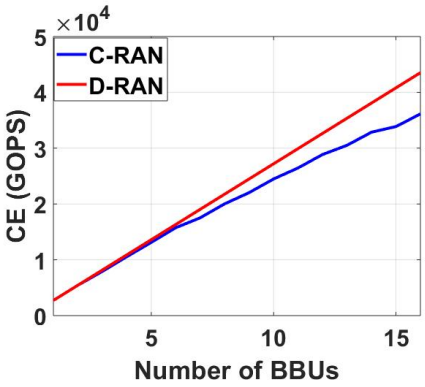


Figure 5.2: Computational effort for different pools dimensions

terminates and aggregates all passing traffic, included traffic which is not destined to the hosted BBUs (i.e., all incoming lightpaths are terminated and processed by the switch). Hence, the main latency contributors considered in this work are the fiber propagation latency and the electronic switch latency. For the sake of simplicity, we assume that every RRH-BBU connection (the so-called "fronthaul" connection) is transported over a dedicated wavelength. In this work we assume that there is a controller at each BBU pool and a centralized C-RAN controller on the top of them [**?**]. Controllers at the pools are responsible for detecting link and BBU failures. When a failure (i.e., BBU failure and/or link failure) is detected, the controller at pool report the C-RAN controller that can activate the backup BBU if needed and reconfigure the network.

The architecture of a node hosting BBU pool is illustrated in Fig. 5.1. An electronic switch with WDM transport interfaces, Gigabit Ethernet (GE) interfaces and Common Public Radio Interfaces (CPRI) is depicted. The arrived lightpaths are multiplexed and demultiplexed by wavelength routers then forwarded to the switch via the WDM

interfaces. Backhaul traffic destined to a hosted BBU is extracted and sent over the GE port, while fronthaul is collected by the switch from the CPRI interfaces, to be mapped into one or more output lightpaths.

5.3.2 Baseband pool computational processing

In this work, we consider the analytical model in [15] to calculate the computational processing of the BBU pool. This model is based on the complexity values estimated by the number of GOPS (what we call "Computational Effort" (CE) in this paper) for the baseband processing functions. We evaluate the computational effort through the following steps. 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 cell sites according to a realistic user-cell site association strategy. After that, we apply a scheduling algorithm to distribute the physical resource blocks of the cell site 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. Fig. 5.2 shows the required CE for the different pool dimensions and compare it to the CE needed by the same number of BBUs in the D-RAN case . In the D-RAN case, the CE scales linearly with number of BBUs, while in C-RAN, the higher number of BBUs per pool the higher savings in computational resources compared to D-RAN.

5.4 Protection Scenarios

To solve the survivable BBU placement problem we consider three different protection approaches as follows:

I) **Dedicated Path Protection (DPP)**: This protection scheme provides resilience against link failure only, as shown in Fig. 3a. Consider BBU of the cell site at node "A" is hosted by the pool at node "D" and routed over the path "AB-BD" as a primary path. In case of link failure, the connection will be routed over the backup path "AE-EF-FD". Note the link disjointness between the two paths.

II) **Dedicated BBU Protection (DBP)**: The second protection scheme provides resilience against BBU failure only, as shown in Fig. 3b. Consider BBU of the cell site at node "A" is hosted

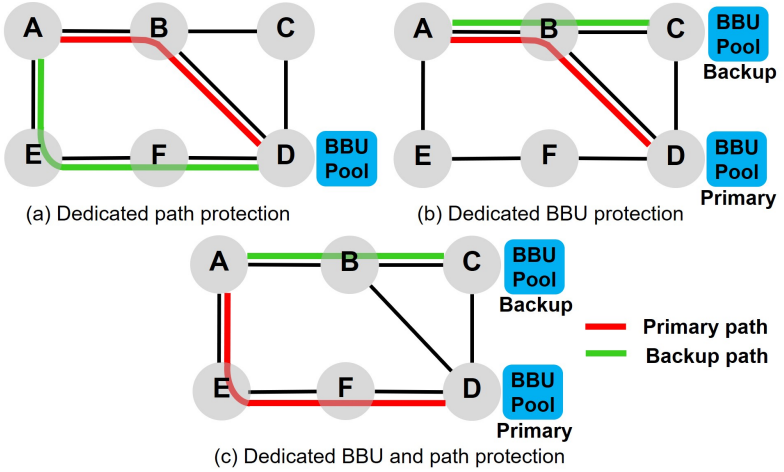


Figure 5.3: Examples of the protection approaches

by the pool at node "D" and routed over the path "AB-BD" as a primary path. In case of BBU failure the connection will be routed to the pool at node "C" through the path "AB-BC". In this approach, path disjointness between the primary and the backup pools is not considered. This protection scheme suits operators' needs when failures occur in nodes with higher probability with respect to links.

III) **Dedicated BBU and Path Protection (DBPP)**: The last protection scheme provides resilience against BBU and link failures, as shown in Fig. 3c. Consider BBU of the cell site at node "A" is hosted by the pool at node "D" and routed over the path "AE-EF-FD" as a primary path. In case of BBU failure and/or link failure the connection will be routed to the pool "C" through the link disjointed path "AB-BC". This protection scheme suits dense urban areas to ensure high availability since BBU outage and/or link failure in those high density areas cause service interruption to high number of users.

5.5 Problem Formulation

In this section we develop ILP formulations for DPP (ILP1), DBP (ILP2) and DBPP (ILP3) protection scenarios. We use a two-layer flow formulation, where an upper virtual layer is made up of virtual links, representing lightpaths originating and terminating in the nodes,

and a lower layer consists of multi-fiber fronthaul links interconnecting the nodes. The survivable BBU pool placement problem is defined as follows:

- **Given**: network topology, number of wavelengths per link, maximum allowed fronthaul latency, and computational effort needed by a pool serving a given number of cell sites.
- **ILP1 output**: placement of BBUs such that each RRH is connected to one BBU pool through two link disjoint paths, one as a primary path and one as backup path.
- **ILP2 output**: placement of BBUs such that each RRH is connected to two BBU pools, one as a primary pool and one as a backup pool.
- **ILP3 output**: placement of BBUs such that each RRH is connected to one primary pool through primary path and one backup pool through backup path where the primary and backup paths are link disjoint.
- **Objective**: minimizing i) number of BBU pools, ii) number of wavelengths and iii) overall computational effort of all BBU pools.

In the following we provide details of the three ILP formulations:

(A) Input sets and parameters:

a) N is set of nodes in the physical network, $i, j, m, n \in N$.

b) E_p is set of physical links, $ij \in E_p$.

c) E_v is set of virtual links, $mn \in E_v$.

d) d_{ij} is the propagation delay introduced by the physical link ij.

e) d_{es} is the latency introduced by the electronic switch at each node.

f) D is the maximum allowable delay between cell site and the BBU pool (fronthaul delay).

g) W is the number of wavelengths per each physical link.

h) C_q is the computational effort in GOPS needed by a pool if it serves q RRHs (the values of C_q are given in Fig. 5.2).

i) C is the maximum computational effort in GOPS that can be accommodated by a pool. Note that at most one pool can be accommodated in each node.

j) M is a large number.

(B) Decision Variables:

a) $k_i = 1$, if node $i \in N$ hosts a BBU pool (binary).

b) $b_{i,q} = 1$, if the BBU pool hosted by node $i \in N$ serves q RRHs as primary and backup (binary). If node i does not host a pool, then $b_{i,0} = 1$.

c) $a_{i,m} = 1$, if cell site at node $m \in N$ is assigned to a BBU pool at node $i \in N$ (binary) (employed only for ILP1).

d) $f_{ij}^{mn} = 1$, if virtual link $mn \in E_v$ between the cell site at node $m \in N$ and the pool at node $n \in N$ is routed over the physical link $ij \in E_p$ as a primary path (binary) (employed only for ILP1).

e) $h_{ij}^{mn} = 1$, if virtual link $mn \in E_v$ between the cell site at node $m \in N$ and the pool at node $n \in N$ is routed over the physical link $ij \in E_p$ as a backup path (binary) (employed only for ILP1).

f) $x_{i,m} = 1$, if cell site at node $m \in N$ is assigned to a primary BBU pool at node $i \in N$ (binary) (employed only for ILP2 and ILP3).

g) $z_{j,m} = 1$, if cell site $m \in N$ is assigned to a backup BBU pool at node $j \in N$ as a backup pool (binary) (employed only for ILP2 and ILP3).

h) $y_{ij}^{mn} = 1$, if virtual link $mn \in E_v$ between the cell site at node $m \in N$ and the primary pool at node $n \in N$ is routed over the physical link $ij \in E_p$ (binary) (employed only for ILP2 and ILP3).

i) $t_{ij}^{mn} = 1$, if virtual link $mn \in E_v$ between the cell site at node $m \in N$ and the backup pool at node $n \in N$ is routed over the physical link $ij \in E_p$ (binary) (employed only for ILP2 and ILP3).

(C) Objective function:

The multi-objective functions illustrated in Eqs. (5.1) and (5.2) are composed of three parts. The First term aims at minimizing the number of BBU pools. Second term minimizes the number

of used wavelengths in the transport network. The last term minimizes the total computational effort required by the network.

I) ILP1

$$min\left(\alpha\sum_{i}k_i+\beta\sum_{ij}\sum_{mn}\left(f_{ij}^{mn}+h_{ij}^{mn}\right)+\gamma\sum_{i}\sum_{q}b_{i,q}\,C_q\right) \tag{5.1}$$

II) ILP2 and ILP3

$$min\left(\alpha\sum_{i}k_i+\beta\sum_{ij}\sum_{mn}\left(y_{ij}^{mn}+t_{ij}^{mn}\right)+\gamma\sum_{i}\sum_{q}b_{i,q}\,C_q\right) \tag{5.2}$$

Parameters α, β and $\gamma \in [0,1]$ can be tuned to select the primary objective of the optimization.

(D) Constraints:

I) ILP1

$$\sum_{i}a_{i,m}=1,\quad \forall m\in N \tag{5.3}$$

$$k_i\geq\frac{\sum_m a_{i,m}}{M},\quad \forall m\in N \tag{5.4}$$

$$\sum_{q}b_{i,q}\,C_q\leq C,\quad \forall i\in N \tag{5.5}$$

$$\sum_{q}q\,b_{i,q}=\sum_{m}a_{i,m},\quad \forall i\in N \tag{5.6}$$

$$\sum_{q}b_{i,q}=1,\quad \forall i\in N \tag{5.7}$$

$$\sum_{mn}\left(f_{ij}^{mn}+h_{ij}^{mn}\right)\leq W,\quad \forall ij\in E_p \tag{5.8}$$

$$\sum_{ij}f_{ij}^{mn}\,(d_{ij}+d_{es})+d_{es}\,(1-x_{m,m})\leq D,\quad \forall mn\in E_v,\forall m\in N \tag{5.9}$$

$$\sum_{ij}h_{ij}^{mn}\,(d_{ij}+d_{es})+d_{es}\,(1-x_{m,m})\leq D,\quad \forall mn\in E_v,\forall m\in N \tag{5.10}$$

$$\sum_j \left(f_{ij}^{mn} - f_{ji}^{mn}\right) = \begin{cases} a_{n,i}, & \text{if } i = m,\ m \neq n. \\ -a_{i,m}, & \text{if } i = n,\ m \neq n. \\ 0, & \text{otherwise.} \end{cases} \quad \forall mn \in E_v, \forall i \in N \tag{5.11}$$

$$\sum_j \left(h_{ij}^{mn} - h_{ji}^{mn}\right) = \begin{cases} a_{n,i}, & \text{if } i = m,\ m \neq n. \\ -a_{i,m}, & \text{if } i = n,\ m \neq n. \\ 0, & \text{otherwise.} \end{cases} \quad \forall mn \in E_v, \forall i \in N \tag{5.12}$$

$$f_{ij}^{mn} + h_{ij}^{mn} \leq 1, \quad \forall mn \in E_v,\ \forall ij \in E_p \tag{5.13}$$

Equation (5.3) enforces that each RRH is associated with exactly one BBU. Equation (5.4) is needed to identify BBU pools as the nodes which host at least one BBU. Equation (5.5) guarantees that GOPS of all the BBUs aggregated in certain pool does not exceed the maximum computational effort for that pool. Equations (5.6) and (5.7) are used to identify what is the computational complexity for the specific number of RRHs served by BBU pool i. Equation (5.8) guarantees that capacity of virtual links routed over a certain physical link does not exceed its capacity. Equations (5.9) and (5.10) ensure the latency requirements for the primary and backup paths, respectively.We take into account both the propagation delay (d_{ij}) and the delay of the electronic switches (d_{es}), whose value is equal to the number of traversed links, plus one. The factor $(1 - x_{m,m})$ is necessary in order to disable this constraint in case the BBU is located at its cell site. Equations (5.11) and (5.12) are the flow constraints, which guarantee that all virtual links are mapped on a set of physical links for primary and backup paths, respectively. Equation (5.13) ensures the disjointness of the primary and backup paths.

II) ILP2

$$\sum_i x_{i,m} = 1, \quad \forall m \in N \tag{5.14}$$

$$\sum_i z_{i,m} = 1, \quad \forall m \in N \tag{5.15}$$

$$x_{i,m} + z_{i,m} \leq 1, \quad \forall i \in N, \forall m \in N \tag{5.16}$$

$$k_i \geq \frac{\sum_m (x_{i,m} + z_{i,m})}{M}, \quad \forall i \in N \tag{5.17}$$

$$\sum_{q} b_{i,q}\, C_q \leq C, \quad \forall i \in N \tag{5.18}$$

$$\sum_{q} q\, b_{i,q} = \sum_{m} (x_{i,m} + z_{i,m}), \quad \forall i \in N \tag{5.19}$$

$$\sum_{q} b_{i,q} = 1, \quad \forall i \in N \tag{5.20}$$

$$\sum_{mn} \left(y_{ij}^{mn} + t_{ij}^{mn}\right) \leq W, \quad \forall ij \in E_p \tag{5.21}$$

$$\sum_{ij} y_{ij}^{mn}\ (d_{ij} + d_{es}) + d_{es}\,(1 - x_{m,m}) \leq D, \quad \forall mn \in E_v, \forall m \in N \tag{5.22}$$

$$\sum_{ij} t_{ij}^{mn}\ (d_{ij} + d_{es}) + d_{es}\,(1 - x_{m,m}) \leq D, \quad \forall mn \in E_v, \forall m \in N \tag{5.23}$$

$$\sum_{j} \left(y_{ij}^{mn} - y_{ji}^{mn}\right) = \begin{cases} x_{n,i}, & \text{if } i = m,\ m \neq n. \\ -x_{i,m}, & \text{if } i = n,\ m \neq n. \\ 0, & \text{otherwise.} \end{cases} \quad \forall mn \in E_v, \forall i \in N \tag{5.24}$$

$$\sum_{j} \left(t_{ij}^{mn} - t_{ji}^{mn}\right) = \begin{cases} z_{n,i}, & \text{if } i = m,\ m \neq n. \\ -z_{i,m}, & \text{if } i = n,\ m \neq n. \\ 0, & \text{otherwise.} \end{cases} \quad \forall mn \in E_v, \forall i \in N \tag{5.25}$$

Equations (5.14) and (5.15) enforce that each RRH is associated with exactly one primary BBU pool and one backup BBU pool. Equation (5.16) enforces primary BBU pool and backup BBU pool to be at different nodes for the same RRH. Equation (5.17) is needed to identify BBU pools as the nodes which host at least one (primary or backup) BBU. Equation (5.18) guarantees that GOPS of all the BBUs aggregated in certain pool does not exceed the maximum computational effort for that pool. Equations (5.19) and (5.20) ensure that the number of RRHs served by pool i equals the sum of RRHs assigned to that pool i. Equation (5.21) guarantees that capacity of virtual links routed over a certain physical link does not exceed its capacity. Equations (5.22) and (5.23) ensure the latency requirements for the primary path and the backup path, respectively (similar to Eq. (5.9) and Eq. (5.10)). Equations (5.24) and (5.25) are the flow constraints, which guarantee that all virtual links connecting the RRHs with the primary BBU pools

and backup BBU pools are mapped on a set of physical links.

III) ILP3

Consider all the constraints for ILP2 with applying the link disjointness between the path to the primary BBU pool and the path to the backup BBU pool as in Eq. (5.26). The cell site at node m can use the physical link ij only once (as a maximum) to route its traffic to any other node n (where n is a potential BBU pool). By this constraint, we can make sure that the physical link ij is not used by the two outgoing paths of node m (the first one to the primary BBU and the second one to the backup BBU) at the same time. As shown in Fig. 3c, the cell site at node "A" routes its traffic to two different BBU pools (pool at node "C" as a primary pool and pool at node "D" as a backup pool). For example, by applying Eq. (5.26), over physical link "BC": $m = A$, $ij = BC$, $n = B, C, E, F, D$, this guarantees that only $y_{BC}^{AC} = 1$. This means that the cell site at node "A" will use the physical link "BC" only once to route the traffic to the primary pool at node "C". Therefore, it (physical link "BC") will not be used to route the traffic of cell site at node "A" to backup pool at node "D" (link disjointness).

$$\sum_{n} \left(y_{ij}^{mn} + t_{ij}^{mn} \right) \leq 1, \quad \forall m \in N, \forall ij \in E_p \tag{5.26}$$

(E) Problem complexity

The total number of variables N_{var} for all three optimization problems is obtained with the formula:

$$N_{var} = |N|(1 + 3|N| + 2|E_p||E_v|) \tag{5.27}$$

The number of constraints varies according to the scenario. In the case of DPP scenario, the number of constraints is given as:

$$N_{DPPconst} = |N|(5 + 4|E_v|) + |E_p|(1 + |E_v|) \tag{5.28}$$

Differently, in the DBP scenario, constraints complexity is:

$$N_{DBPconst} = |N|(6 + |N| + 4|E_v|) + |E_p| \tag{5.29}$$

Finally, in the DBPP scenario, constraints complexity is:

$$N_{DBPPconst} = |N|(6 + |N| + 4|E_v| + |E_p|) + |E_p| \tag{5.30}$$

For the number of variables N_{var}, we observe that the dominate term is $2\,.\,|N|\,.\,|E_p|\,.\,|E_v|$, while for the constraints complexity (for the different scenarios) the dominate term is $4\,.\,|N|\,.\,|E_v|$. Therefore, the problem complexity for all the proposed protection scenario, given by the sum of the number of variables and the number of constraints, is in the order of $O(|N|\,.\,|E_v|\,.\,|E_p|)$.

5.6 Illustrative Numerical Results

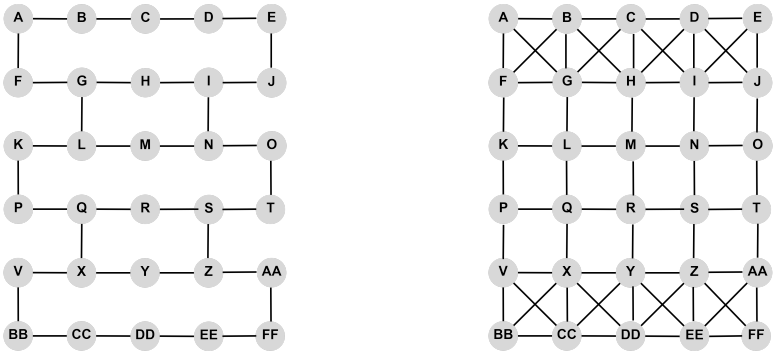
5.6.1 Evaluation Settings

In this work, we consider two different aggregation network topologies with different levels of node connectivity degree, as in [62] (i.e., ratio between the total number of edges and the total number of nodes in the network) as shown in Fig. 5.4. Each topology consists of 30 nodes uniformly distributed over a dense urban square region of 7.5 km^2. For simplicity, we assume that every node contains one RRH. Every node represents a potential BBU pool. We consider a 20 μs latency for each electronic switch and one fiber per link, carrying $W = 6$ wavelengths at 10 Gb/s. We assume that the pool can accommodate 20 BBUs as maximum with 44080 GOPS. To solve our optimizations we used ILOG CPLEX 12.0 on a workstation equipped with 8 $\times$ 2 GHz processors and 32 GB of RAM. In this section,we show the results for the following objective functions.

- **First objective function (O1)**: we optimize the number of BBU pools as a first objective, then the number of used wavelengths as the second objective.
- **Second objective function (O2)**: we first optimize the number of used wavelengths then the number of BBU pools.
- **Third objective function (O3)**: we optimize first the computational effort, then the number of BBU pools, then the number

Table 5.1: Parameters to select the objective function

	α	β	γ
O1	1	10^{-3}	0
O2	10^{-3}	1	0
O3	1	10^{-3}	10



(a) Topology "a" (b) Topology "b"

Figure 5.4: Network Topologies

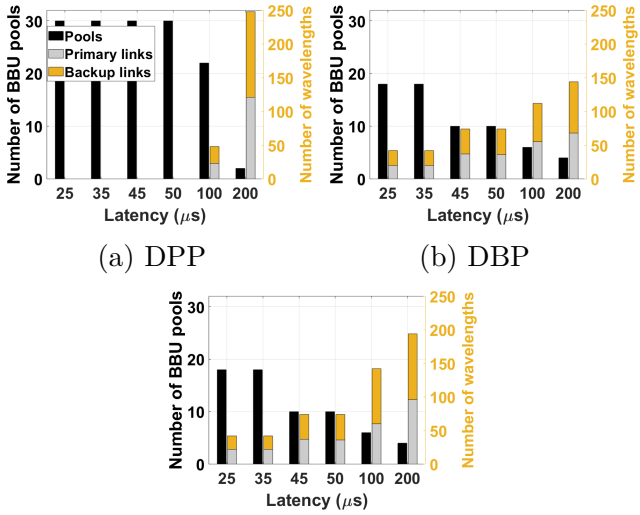
of used wavelengths.

The values of the parameters α, β, γ to select the desired objective function are given in Table 1. Note that the number of wavelengths is counted as the summation of the variables y_{ij}^{mn} and t_{ij}^{mn} in DBP and DBPP scenarios and as summation of the variables f_{ij}^{mn} and h_{ij}^{mn} in DPP scenario.

5.6.2 Numerical Results

Fig. 5.5 shows the number of BBU pools and the number of used wavelengths (left and right bars, respectively) as a function of the maximum allowable fronthaul latency (D) considering the three protection approaches and objetive function O1 in network topology "a" (connectivity degree 1.13).

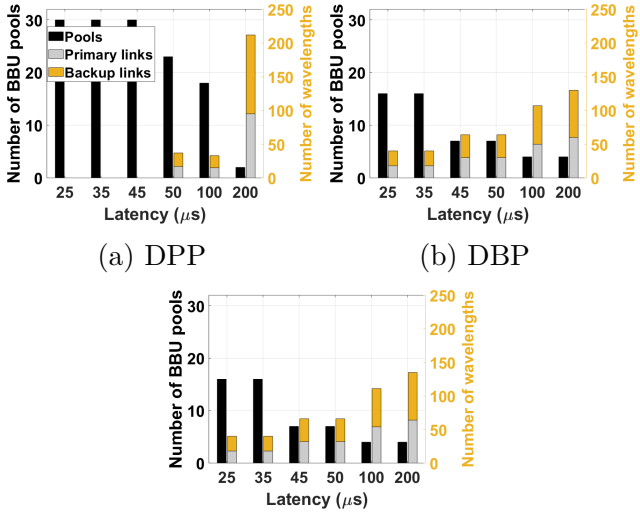
For DPP (Fig. 5.5a) at low latency values ($D \leq 50\,\mu s$) no centralization occurs. All BBUs are placed at the cell sites (no wavelengths utilized for the transport network) as the tight latency constraint does not allow the BBUs to be separated from the RRHs and to interconnect via two link-disjoint paths. To this end, the solution obtained is to distribute all the BBUs (place each BBU at its cell site); hence no links are established to perform link protection. As the latency value increases ($D = 100\,\mu s$) only few BBUs can be centralized with the two disjoint paths resulting in high number of BBU pools (22 BBU pools) and low number of utilized wavelengths (48 wavelengths). For more relaxed latency values ($D = 200\,\mu s$), the number of BBU pools decreases (2 pools) while the number of wavelengths increases (250 wavelengths). Note that, the minimum number of pools can be



(a) DPP (b) DBP

(c) DBPP

FIGURE 5.5: Number of BBU pools and wavelengths (Topology = "a", Objective function = O1)



(a) DPP (b) DBP

(c) DBPP

FIGURE 5.6: Number of BBU pools and wavelengths (Topology = "b", Objective function = O1)

achieved is 2 as one pool can accommodate up to 20 BBUs and the number RRHs in the network is 30.

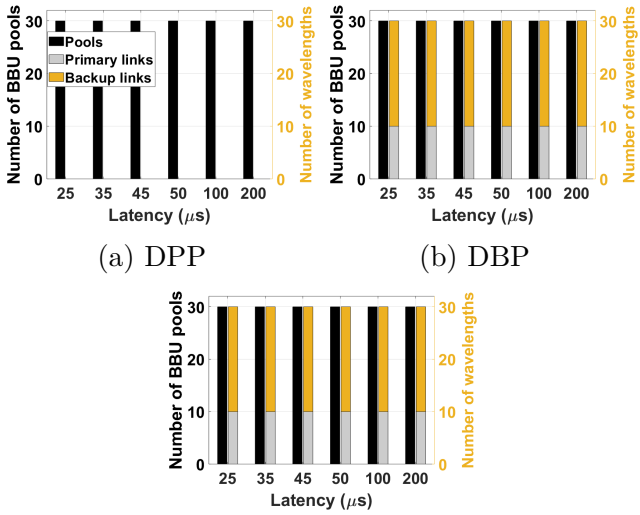
For DBP and DBPP (Fig. 5.5b and Fig. 5.5c, respectively), no solution could be obtained for $D < 25\,\mu s$, as the strict latency does not

allow BBUs to be hosted outside the cell site; hence no BBU protection can be guaranteed in these cases. For low latency values ($D \leq 45\,\mu s$) most of the nodes (18 nodes) must be activated as pools. In this case, the tight latency values force one BBU (the primary or the backup) to be placed at the cell site while the other BBU is placed in one of the adjacent nodes ,resulting in low number of utilized wavelengths. As the maximum allowable latency increases, the number of BBU pools decreased until it reaches 4 pools at $D = 200\,\mu s$ with the maximum number of utilized wavelengths.

By comparing the three protection approaches, DBP and DBPP have the same performance except for high latency values ($D \geq 100\,\mu s$), the number of utilized wavelengths in DBPP is higher by 35% as in DBPP path protection is considered. DPP does not provide centralization nor protection at tight latency values ($D < 100\,\mu s$). For latency value $D = 100\,\mu s$, DPP can be guaranteed with higher number of BBU pools (22 pools) compared to the two other scenarios (6 pools) although it does not provide BBU protection. This is due to the tight latency constraint on both the primary and backup paths to reach the same pool, forcing high number of BBUs to be placed at their cell sites; hence high number of nodes are activated as a BBU pool. One the other hand, in DBP and DBPP, meeting those latency values is applicable -with higher degree of centralization compared to DPP- as we route the connections to two disjoint pools (primary and backup), leading to less number of the pools. At high latency value $D = 200\,\mu s$, DPP has low number of BBU pools (2 pools) compared to DBP and DBPP (4 pools) as BBU protection is not considered in DPP. Subsequently, the number of utilized wavelengths in DPP (248 wavelengths) is higher than that in DBP and DBPP (144 and 194 wavelengths, respectively).

Fig. 5.6 shows the number of BBU pools and the number of used wavelengths (left and right bars, respectively) as a function of the maximum allowable fronthaul latency (D) considering the three different protection approaches and objective function O1 in network topology “b” (connectivity 2.16). It is clear that the results for topology “b” follow the same manner of topology “a”. Unlike toplogy “a”, DBP (Fig. 5.6b) and DBPP (Fig. 5.6c) have almost equal number wavelengths at all the latency values. This can be explained considering that as the connectivity is higher compared to topology “a”, the disjointness of the primary and backup paths does not contribute to the number of utilized wavelengths.

Fig. 5.7 shows the number of BBU pools and wavelengths as a function of the maximum allowable fronthaul latency (D) considering



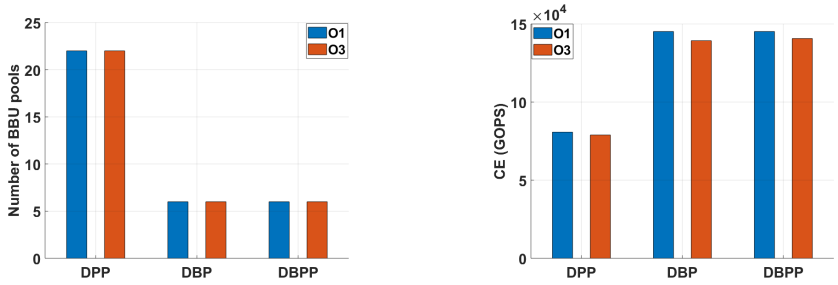
(a) DPP (b) DBP

(c) DBPP

FIGURE 5.7: Number of BBU pools and wavelengths (Topology = "a", Objective function = O2)

the three protection approaches and objective function O2 in network topology "a". For DPP (Fig. 5.7a), no centralization occurs, the BBUs are placed at the cell sites for all the latency values. This can be explained by that as minimizing the number of wavelengths is the first priority in O2 so the BBUs are placed at the cell site leading to zero utilized wavelengths, subsequently, no RRH-BBU connections to protect. For DBP (Fig. 5.7b) and DBPP (Fig. 5.7c), the number of BBU pools and number of used wavelengths are constant over the different latencies values. As the first objective of O2 is to minimize the number of wavelengths, the solver keeps one BBU (the primary or the backup) at the cell site and allocates the other BBU in the adjacent node leading to minimum as well as constant number of wavelengths. Note that we obtain the same results for topology "b", since in DPP all BBU located at the cell site so connectivity does not contribute. Similarly, for DBP and DBPP the BBU is placed at the adjacent node.

Fig. 5.8 compares objectives O1 and O3 in terms of the number of pools and the amount of computational effort considering topology "a" for $D = 100\,\mu s$. Fig. 5.8a shows the number of BBU pools for O1 and O3 considering the three protection approaches. The results show that adding a term to minimize the computational effort does not contribute to the number of the BBU pools as no significant difference is observed between O1 and O3 for all the protection approaches. Fig. 5.8b shows the computational effort in GOPS corresponding to each



(a) Number of BBU pools Vs protection approach

(b) Computational effort in GOPS Vs protection approach

Figure 5.8: Number of BBU pools and the computational effort (Topology = "a", Objective function = O1 and O3)

objective function and protection approach. Despite O1 and O3 have the same number of BBU pools, the two objectives provide different amount of computational effort. While minimizing the GOPS in O3 not only guarantee the minimum number of pools but also choose combinations for the number of BBUs to be placed in the pools with higher multiplexing gain (less computational effort). By using O3 we estimate saving in computational effort up to 2.2% in DPP, 4.1% in DBP and 3.1% in DBPP with respect to O1. In general, for other network topologies, minimizing the computational effort might not give the minimum number of BBU pools although giving the lowest amount of GOPS.

5.7 Conclusion

In this work, we have investigated the survivable C-RAN deployment problem over an optical aggregation network through three protection approaches namely: i) dedicated path protection ii) dedicated BBU protection iii) dedicated BBUand path protection. We have formulated optimization problem for each approach through an ILP model with different objectives, minimizing number of BBU pools, number of used wavelengths and overall computational effort. We show the results of the optimization problems for the three approaches on two different network topologies with different connectivity values considering latency, link capacity and the BBU pool computational resource capacity constraints. Finally, we show that minimizing the computational effort results in additional savings compared to the traditional minimization

of the number of BBU pools.

As a future work, we will consider grooming and wavelength assignment features in the model. Moreover, we plan to develop a heuristic algorithm to deal with larger network instances.

BBU Placement with Shared Path Protection Schemes in 5G C-RAN over Optical Aggregation Networks 6

In this chapter, the survivable BBU placement problem introduced in the previous chapter is evolved towards shared protection direction. We introduce a BBU placement problem with shared protection approach to have more efficient wavelengths utilization.

6.1 Introduction

In the previous chapter, we introduced an optimization problem for placing the BBUs in a WDM aggregation network with different dedicated protection. In this chapter, we continue investigating the survivable C-RAN different protection schemes shifting our focus to shared protection approach. We introduce two approaches for a shared path protection with wavelength assignment. The first approach is dedicated BBU protection and shared path protection with wavelength continuity. While the second approach is dedicated BBU protection and shared path protection without wavelength continuity. We formalize the approaches as an ILP problem and solve it over a 5G optical aggregation network. We investigate the effect of wavelength continuity on the latency contribution. Moreover, we show the efficient utilization of wavelengths of the shared protection scheme compared to the dedicated one.

The remainder of the chapter is organized as follows. Section 6.2

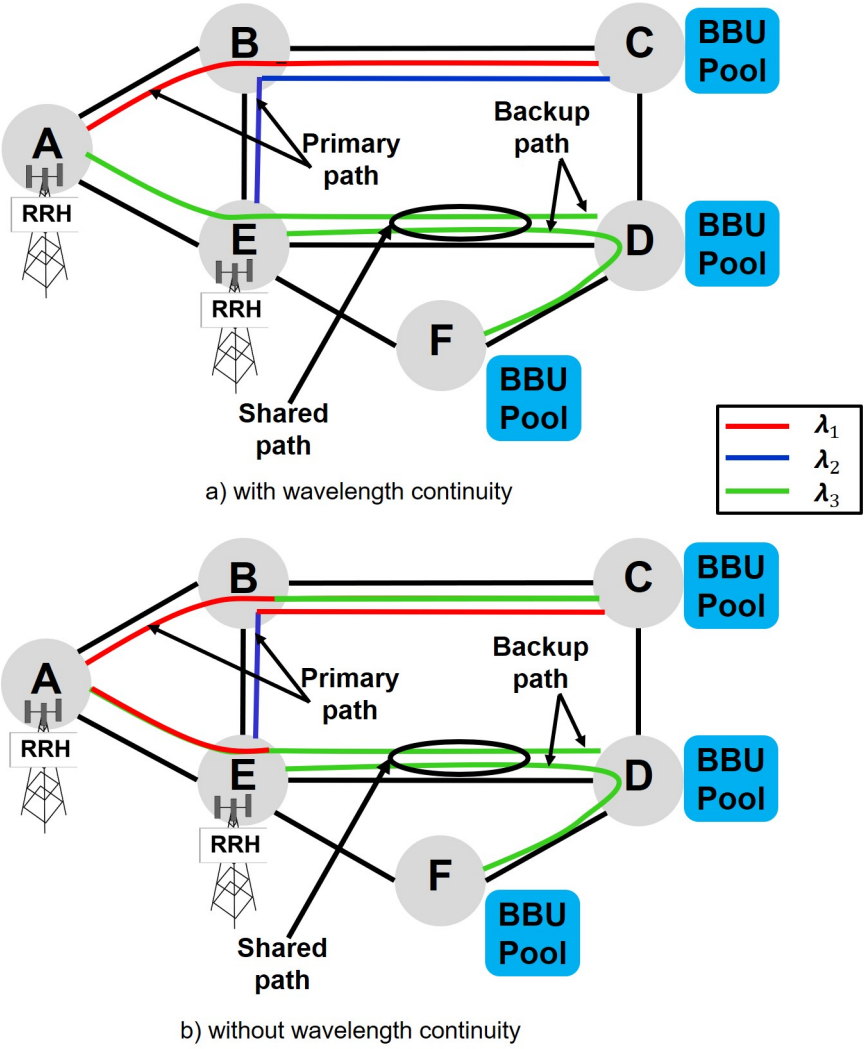


Figure 6.1: Shared protection approaches

discusses the considered protection approaches. Section 6.3 shows the ILP formulations used to solve the resilient problem. Section 6.4 shows the illustrative numerical results. Section 6.5 concludes the work

6.2 Protection Scenario

In this section, we introduce a survivable BBU placement problem considering dedicated BBU protection and shared path protection. We consider two different approcahes i) with wavelength continuity ii) without wavelength continuity.

I) **Dedicated BBU Shared Path Protection with Wavelength Continuity:** We jointly consider protection against BBU and path failures, as shown in Fig. 6.1a. Consider BBU of the cell site at node "A" is hosted by the pool at node "C" and routed through the path "AB-BC" over wavelength λ_1 as a primary path. In case of BBU failure and/or link failure the connection will be routed to the pool "D" through the disjointed path "AE-ED" over wavelength "λ_3". Consider BBU of the cell site at node "E" is hosted by the pool at node "C" and routed over the path "EB-BC" and wavelength "λ_2" as a primary path. In case of BBU failure and/or link failure the connection will be routed to the pool "F" through the disjointed path "ED-DF" over wavelength "λ_3". The previous mentioned scenario illustrates the following: 1) Wavelength continuity is considered in primary and backup paths. 2) The primary connection (e.g., "AC" and "EC") is routed over dedicated paths. 3) The secondary paths can share wavelength over some links("λ_3" in link "ED").

II) **Dedicated BBU and Shared Path Protection without Wavelength Continuity:** We apply the same previous approach without considering wavelength continuity over the paths as shown in Fig. 6.1b. For example, the BBU of the cell site at node "E" is hosted by the pool at node "C" and routed through the path "EB-BC" over two different wavelengths "λ_2" and "λ_1". We add in each node a wavelength converter so in this approach we count two latency contributors: propagation delay and wavelength converter latency.

6.3 Problem Formulation

In this section we develop ILP formulations for Dedicated BBU Shared Path Protection with Wavelength Continuity (ILP1) and Dedicated BBU Shared Path Protection without Wavelength Continuity (ILP2) protection scenarios. We use a two-layer flow formulation, where an upper virtual layer is made up of virtual links, representing lightpaths originating and terminating in the nodes, and a lower layer consists of multi-fiber fronthaul links interconnecting the nodes. The survivable BBU pool placement problem is defined as follows:

- **Given**: network topology, set of wavelengths, maximum allowed fronthaul latency, and computational effort needed by a pool serving given number of cell sites.

- **Output**: placement of BBUs such that each RRH is connected to one primary pool through primary path and wavelength and one backup pool through backup path and wavelength where the primary and backup paths are disjoint, and routing of traffic between the RRHs to the BBU pool.
- **Objective**: minimizing i) number of BBU pools and ii) number of wavelengths

In the following we provide details of the ILP formulation:

1. Input sets and parameters:
   a) N is set of nodes in the physical network, $i, j, m, n \in N$.
   b) W is set of wavelengths , $\lambda \in W$.
   c) d_{ij} is the propagation delay introduced by the physical link i-j.
   d) d_c delay introduced by the wavelength converter (employed only for ILP2)
   e) D is the maximum allowable delay between cell site and the BBU pool (fronthaul delay).
   f) C_q is the computational effort in GOPS needed by a pool if it serves q RRHs (the values of C_q are given in Fig. 5.2).
   g) C is the maximum computational effort in GOPS that can be accommodated by a pool. Note that at most one pool can be accommodated in each node.
   h) S is the maximum number of virtual links routed over a certain wavelength of a physical link (sharing factor).
   i) M is big number.
2. Decision Variables:
   a) $k_i = 1$, if node i hosts a BBU pool (binary).
   b) $b_{i,q} = 1$, if the BBU pool hosted by node i serves q RRHs as primary and backup (binary). If node i does not host a pool, then $b_{i,0} = 1$.
   c) $x_{i,m} = 1$, if cell site m is assigned to a primary BBU pool at node i (binary).
   d) $z_{j,m} = 1$, if cell site m is assigned to a backup BBU pool at node j as a backup pool (binary).

e) $y_{ij,\lambda}^{mn} = 1$, if virtual link m-n to reach a primary pool is routed over the physical link i-j and wavelength λ (binary).

f) $t_{ij,\lambda}^{mn} = 1$, if virtual link m-n to reach a backup pool is routed over the physical link i-j and wavelength λ (binary).

g) $v_{\lambda}^{mn} = 1$, if virtual link m-n is routed over wavelength λ (employed only for ILP1) (binary).

3. Objective function:

   The multi-objective function illustrated in Eq. (6.1) is composed of two parts. First term aims at minimizing the number of BBU pools. Second term minimizes the number of used wavelengths in the transport network.

$$min \left(\alpha \sum_{i} k_i + \alpha \sum_{ij} \sum_{mn} \sum_{\lambda} \left(y_{ij,\lambda}^{mn} + t_{ij,\lambda}^{mn} \right) \right) \tag{6.1}$$

   Parameters α and $\beta \in [0,1]$ can be tuned to select the primary objective of the optimization.

4. Constraints:

   I) ILP1

$$\sum_{i} x_{i,m} = 1, \quad \forall m \tag{6.2}$$

$$\sum_{i} z_{i,m} = 1, \quad \forall m \tag{6.3}$$

$$x_{i,m} + z_{i,m} \leq 1, \quad \forall i, \forall m \tag{6.4}$$

$$k_i \geq \frac{\sum_m x_{i,m} + z_{i,m}}{M}, \quad \forall m \tag{6.5}$$

$$\sum_{q} b_{i,q} C_q \leq C, \quad \forall i \tag{6.6}$$

$$\sum_{q} q\, b_{i,q} = \sum_{m} x_{i,m} + z_{i,m}, \quad \forall i \tag{6.7}$$

$$\sum_{q} b_{i,q} = 1, \quad \forall i \tag{6.8}$$

$$\sum_{ij} y_{ij,\lambda}^{mn} \, d_{ij} \leq D, \quad \forall mn \tag{6.9}$$

$$\sum_{ij} t_{ij,\lambda}^{mn} \, d_{ij} \leq D, \quad \forall mn \tag{6.10}$$

$$\sum_{j} \left(y_{ij,\lambda}^{mn} - y_{ji,\lambda}^{mn} \right) = \begin{cases} x_{n,i}, & \text{if } i = m,\ m \neq n. \\ -x_{i,m}, & \text{if } i = n,\ m \neq n. \\ 0, & \text{otherwise.} \end{cases} \quad \forall mn, \forall i \tag{6.11}$$

$$\sum_{j} \left(t_{ij,\lambda}^{mn} - t_{ji,\lambda}^{mn} \right) = \begin{cases} z_{n,i}, & \text{if } i = m,\ m \neq n. \\ -z_{i,m}, & \text{if } i = n,\ m \neq n. \\ 0, & \text{otherwise.} \end{cases} \quad \forall mn, \forall i \tag{6.12}$$

$$\sum_{n} y_{ij,\lambda}^{mn} + t_{ij,\lambda}^{mn} \leq 1, \quad \forall m, \forall ij \tag{6.13}$$

$$\sum_{mn} t_{ij,\lambda}^{mn} \leq S, \quad \forall ij, \forall \lambda \tag{6.14}$$

$$\sum_{mn} y_{ij,\lambda}^{mn} + \frac{t_{ij,\lambda}^{mn}}{S} \leq 1, \quad \forall ij, \forall \lambda \tag{6.15}$$

$$\sum_{\lambda} v_{\lambda}^{mn} \leq 1, \quad \forall mn \tag{6.16}$$

$$\sum_{ij} y_{ij,\lambda}^{mn} \leq M \, v_{\lambda}^{mn}, \quad \forall mn, \forall \lambda \tag{6.17}$$

$$\sum_{\text{ij}} t_{ij,\lambda}^{mn} \leq M \, v_{\lambda}^{mn}, \quad \forall mn, \forall \lambda \tag{6.18}$$

Equations (6.2) and (6.3) enforce that each RRH is associated with exactly one primary BBU pool and one backup BBU pool. Equation (6.4) enforces primary BBU pool and backup BBU pool to be at different nodes for the same RRH. Equation (6.5) is needed to identify BBU pools as the nodes which host at least one (primary or backup) BBU. Equation (6.6) to guarantee that GOPS of all the BBUs

aggregated in certain pool does not exceed the maximum computational effort for that pool. Equations (6.7) and (6.8) ensure that the number of RRHs served by pool i equals the sum of RRHs assigned to that pool i. Equations (6.9) and (6.10) ensure the latency requirements. Equations (6.11) and (6.12) are the flow constraints, which guarantee that all virtual links connecting the RRHs with the primary BBU pools and backup BBU pools are mapped on a set of physical links. Equation (6.13) ensures the disjointness between the path to the primary BBU pool and the path to the backup BBU pool. Equation (6.14) ensure the the number of virtual links routed over a certain wavelength of a physical link does not exceed the sharing factor. Equation (6.15) ensures that the wavelength over a certain link used by primary path is not used by any other primary or backup path. Equations (6.16), (6.17) and (6.18) ensure wavelength continuity.

II) ILP2

Consider all the constraints for ILP1 with eliminating those for wavelength continuity (i.e., Eqs. (6.16), (6.17) and (6.18)). We consider two latency contributers the propagation delay and the wavelength converter, hence we replace Eqs. (6.9) and (6.10) by Eqs.

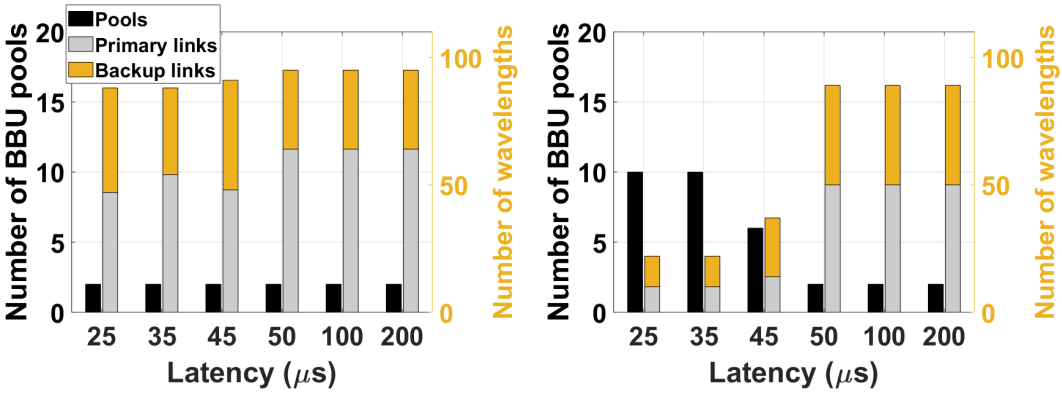
$$\sum_{ij} y_{ij,\lambda}^{mn} \, (d_{ij} + d_c) \leq D, \quad \forall mn \tag{6.19}$$

$$\sum_{ij} t_{ij,\lambda}^{mn} \, (d_{ij} + d_c) \leq D, \quad \forall mn \tag{6.20}$$

6.4 Illustrative Numerical Results

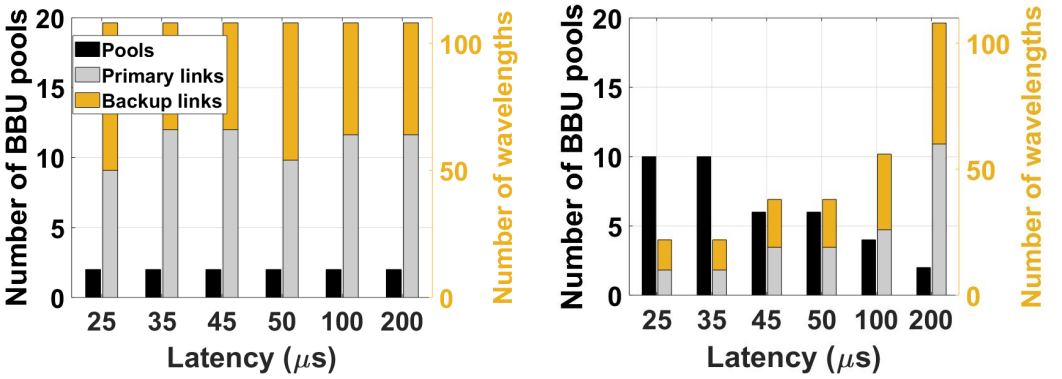
6.4.1 Evaluation Settings

In this chapter, we consider the network topology shown in Fig. **??** which consists of 16 nodes uniformly distributed over a dense urban square region of 4 km^2 . We consider a 20 μs latency for each wavelength converter. We consider the set of wavelengths W consists of 6 different wavelengths. We assume that the pool can accommodate 16 BBUs as maximum with 36128 GOPS. To solve our optimizations we used ILOG CPLEX 12.0 on a workstation equipped with 8 $\times$ 2 GHz



(a) Wavelength continuity (b) Without wavelength continuity

FIGURE 6.2: Number of BBU pools and wavelengths "S=3"



(a) Wavelength continuity (b) Without wavelength continuity

FIGURE 6.3: Number of BBU pools and wavelengths "S=1"

processors and 32 GB of RAM. In this section, we show the results for the two following cases:

- **Case 1:** Sharing factor (S) =3
- **Case 2:** Sharing factor (S) =1 (i.e., no sharing)

6.4.2 Numerical Results

Fig. 6.2 shows the number of BBU pools and the number of used wavelengths (left and right bars, respectively) as a function of the maximum allowable fronthaul latency (D) considering $S = 3$ for the two protection approaches (i.e., with and without wavelength continuity).

When considering wavelength continuity (Fig. 6.2a) full centralization occurs even in low latencies as we only count the propagation delay. The number of used wavelengths increase slightly as D increases

In Fig. 6.2b wavelength continuity is not considered. We count the wavelength converter latency and the propagation delay. For low D values ($D < 45\,\mu s$) most of the nodes must be activated as pools. In this case, the tight latency values force one BBU (the primary or the backup) to be placed at the cell site while the other BBU is placed in one of the adjacent nodes ,resulting in low number of utilized wavelengths. As the maximum allowable latency increases, the number of BBU pools decreased until we achieve full centralization at $D = 50\,\mu s$ with the maximum number of utilized wavelengths.

Comparing shared path protection (Fig. 6.2) to dedicated path protection (Fig. 6.3) we can observe similar behavior except that the number of utilized wavelengths in dedicated protection is more than that for shared protection. It can be easily proven that shared protection has efficient wavelengths utilization comapred to dedicated protection.

6.5 Conclusion

In this chapter we have investigated different shared path protection approaches for C-RAN deployment over optical aggregation network. We formulate two protection schemes namely: i) Dedicated BBU protection and shared path protection with wavelength continuity ii) Dedicated BBU protection and shared path protection without wavelength continuity. We show how considering wavelength continuity can be suitable for low latency application as only propagation delay is considered. While not considering wavelength continuity results in higher latency due to the effect of wavelength converter. Moreover, we compare between shared path protection and dedicated path protection showing the savings in wavelengths can be achieved in the shared one.

Concluding Remarks 7

In this thesis, we have investigated the computational, cost, energy and survivability effectiveness of 5G access network architecture, based on a promising principle referred to as "C-RAN". "C-RAN" is a new radio access network paradigm in which the BaseBand Units are centralized into single physical location and the new fronthaul (CPRI) traffic is introduced into the network.

In Chapter 2, we made a survey of the technological features of C-RAN and implementation principles from a networking standpoint.

In Chapter 3, we present an analytical model to estimate the computational savings (what we called multiplexing gain) enabled by BBUs centralization. By the aid of the previous mentioned model we elaborate a cost model to show the cost reduction due BBU centralization. Moreover, we discuss the restricted fronthaul requirements (in terms of very high bit rate and low latency) attached with that centralization. In order to, mitigate the fronthaul requirements we investigate four different RAN functional splits, i.e., CPRI split, PHY split, PHY-MAC split, RLC-PDCP split. Then, we extend our multiplexing gain model to capture the computational savings in those splits. Results show that up to 28% computational savings can be achieved across the different splits which affects the cost by 30% savings.

After estimating the multiplexing gain, in Chapter 4 we move forward to calculate the power consumption of centralized digital units for the previous mentioned RAN functional splits. We mention the main power consumption contributers, namely: fronthaul interface, S1-X2 interface, processing cards and baseline unit. Our model is based on calculating the GOPS of each baseband function. We compare

the power consumption of the centralized units to the traditional scenario (distributed RAN) the results show that 24% power savings is achievable.

In chapter 5, we investigate the survivability of the C-RAN. The survivability of a network refers to a network's capability to provide continuous service in the presence of failures. Survivability is one of the most important aspects as the BBU failure may cause service outage to huge number of users. We propose three different approaches for the survivable BBU pool placement problem and traffic routing in C-RAN deployment over a 5G optical aggregation network. Namely we define the following protection scenarios: i) Dedicated Path protection, ii) Dedicated BBU protection and iii) Dedicated BBU and path protection. The three approaches are formalized as ILP problems with objectives to minimize the number of BBU pools, the number of used wavelengths and the baseband processing computational resources, in terms of GOPS. We provide numerical results to compare the aforementioned protection strategies and optimization objectives showing the effect of the latency and of the transport network capacity on the BBU placement.

In chapter 6, we continue investigating the survivability of the C-RAN, shifting our focus to shared protection schemes. We propose two different approaches one with considering wavelength continuity and the second without considering continuity. The two approaches are formalized as ILP problems to minimize the number of activated BBU pools and the number of used wavelengths. The results compare the performance of the two approaches at different allowable latency values. Moreover we show the efficient utilization of the available wavelengths for the shared protection over the dedicated protection.

Although the presented research is in a initial stage, we believe that the defined problems are enough general and powerful to serve as basis for many possible extensions, in order to study some interesting open issues that appeared during this work.

For instance, the proposed multiplexing gain, power and survivability models assume a static traffic characterization. A possible extension is about a dynamic traffic scenario, in which the models also embed the statistical information about daily traffic variation patterns (i.e., the "tidal effect"). Within such scenario, it may be interesting to integrate per-node energy efficiency strategies (e.g., ON/OFF, "sleep" or low-power modes) in the network optimization, in order to dynamically adapt the configuration of some nodes to the varying traffic conditions.

Regarding the introduced multiplexing gain and power models, including the transport fronthaul network in the problem would give a

better view on the savings introduced by the C-RAN deployment.

Finally, some considerations regarding the next 5G mobile access can be applied. Even though the 5G standardization process still in progress, there is no doubt that it will enhance the network performance, not only in terms of higher data rates per user and lower latency, but also in terms of "intelligence" of the network. To achieve this, 5G networks are expected to integrate some solution as, cell densification by means of massive small cell deployment (micro-, femto-, etc.), coordinated multi-cell and multi-antenna processing (CoMP, eICIC) and Centralized/Cloud RAN. Such techniques will ultimately burden the wireline part of access networks and the aggregation networks, because around tens of Gb/s of data must be transported, due to backhaul, coordination and fronthaul traffic, with sub-ms latency constraints. As a consequence, an optimized design of the optical access/aggregation networks, capable of satisfying all such requisites, is expected to have an increased importance, since its performance will directly influence the performance of the 5G network.

Bibliography

[1] T. Zhou *et al.*, "Load-aware user association with quality of service support in heterogeneous cellular networks," *IET Communications*, vol. 9, no. 4, pp. 494–500, 2015.

[2] "Ericson Mobility Report," Ericson, Technical Report (TR), Jun. [Online]. Available: https:www.ericsson.com/assets/local/news/2015/6/ericsson-mobility-report-june-2015.pdf

[3] K. Chen and R. Duan, "C-ran the road towards green ran," *China Mobile Research Institute, white paper*, vol. 2, 2011.

[4] L. Belkhir and A. Elmeligi, "Assessing ict global emissions footprint: Trends to 2040 & recommendations," *Journal of Cleaner Production*, vol. 177, pp. 448–463, 2018.

[5] M. Olsson *et al.*, "5GrEEn: Towards Green 5G mobile networks," in *IEEE 9th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*, Lyon, France, Oct. 2013, pp. 212–216.

[6] "5G White Paper," Next generation Mobile Network (NGMN) Alliance, White paper, Feb. 2015. [Online]. Available: https://www.ngmn.org

[7] H. Tullberg *et al.*, "The METIS 5G system concept: Meeting the 5G requirements," in *IEEE Communications Magazine*, vol. 54, no. 12. IEEE, 2016, pp. 132–139.

[8] A. Checko *et al.*, "Cloud RAN for mobile networks- A technology overview," *IEEE Communications surveys & tutorials*, vol. 17, no. 1, pp. 405–426, Mar. 2015.

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

[10] Y. Huiyu *et al.*, "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.

[11] A. Pizzinat *et al.*, "Things you should know about fronthaul," *Journal of Lightwave Technology*, vol. 33, no. 5, pp. 1077–1083, Mar. 2015.

[12] J. P. Sterbenz *et al.*, "Resilience and survivability in communication networks: Strategies, principles, and survey of disciplines," *Computer Networks*, vol. 54, no. 8, pp. 1245–1265, 2010.

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

[14] M. De Andrade *et al.*, "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.

[15] M. Shehata *et al.*, "Multiplexing gain and processing savings of 5g radio-access-network functional splits," *IEEE Transactions on Green Communications and Networking*, 2018.

[16] M. Agiwal, A. Roy, and N. Saxena, "Next generation 5G wireless networks: A comprehensive survey," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 3, pp. 1617–1655, 2016.

[17] K. Ramantas *et al.*, "New unified PON-RAN access architecture for 4G LTE networks," *Journal of Optical Communications and Networking*, vol. 6, no. 10, pp. 890–900, 2014.

[18] J. Liu *et al.*, "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.

[19] 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.

[20] 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.

[21] S. Bhaumik *et al.*, "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.

[22] S. Namba *et al.*, "Colony-RAN architecture for future cellular network," in *Proc. Future Network & Mobile Summit (FutureNetw)*, Berlin, Germany, Jul. 2012, pp. 1–8.

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

[24] A. Maeder *et al.*, "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.

[25] J. Bartelt *et al.*, "Fronthaul and backhaul requirements of flexibly centralized radio access networks," *IEEE Wireless Communications*, vol. 22, no. 5, pp. 105–111, Oct. 2015.

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

[27] M. Jaber *et al.*, "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.

[28] X. Wang *et al.*, "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.

[29] U. Dötsch *et al.*, "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.

[30] L. Valcarenghi *et al.*, "Requirements for 5G fronthaul," in *Proc. 18th International Conference on Transparent Optical Networks (ICTON)*, Trento, Italy, Jul. 2016, pp. 1–5.

[31] K. Miyamoto *et al.*, "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.

[32] J. Liu *et al.*, "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.

[33] "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

[34] "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

[35] 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.

[36] D. Lopez-Perez *et al.*, "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.

[37] 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.

[38] R. Van Nee *et al.*, "The 802.11 n MIMO-OFDM standard for wireless LAN and beyond," *Wireless Personal Communications*, vol. 37, no. 3, pp. 445–453, May 2006.

[39] 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.

[40] "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

[41] "Analysis of Transport Network Architectures for Structural Convergence," COnvergence of fixed and Mobile BrOadband

access/aggregation networks- COMBO Project, Tech. Rep. D3.3, Jul. 2015. [Online]. Available: https://www.ict-combo.eu

[42] G. Auer *et al.*, "How much energy is needed to run a wireless network?" *IEEE Wireless Communications*, vol. 18, no. 5, pp. 40–49, Oct. 2011.

[43] S. Tombaz *et al.*, "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.

[44] 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.

[45] 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.

[46] L. Chen *et al.*, "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.

[47] X. Wang *et al.*, "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.

[48] C. Desset, B. Debaillie, and F. Louagie, "Towards a flexible and future-proof power model for cellular base stations," in *24th Tyrrhenian International Workshop on Digital Communications-Green ICT (TIWDC)*, Genoa, Italy, 2013, pp. 1–6.

[49] 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.

[50] C. Desset *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.

[51] F. Musumeci *et al.*, "Optimal BBU placement for 5G C-RAN deployment over WDM aggregation networks," *Journal of Lightwave Technology*, vol. 34, no. 8, pp. 1963–1970, 2016.

[52] N. Carapellese, M. Tornatore, and A. Pattavina, "Energy-efficient baseband unit placement in a fixed/mobile converged WDM aggregation

network," *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 8, pp. 1542–1551, 2014.

[53] R. Riggio *et al.*, "SWAN: Base-band units placement over reconfigurable wireless front-hauls," in *12th International Conference on Network and Service Management (CNSM)*. Montreal, Canada: IEEE, 2016, pp. 28–36.

[54] S. S. Lisi *et al.*, "Cost-effective migration towards C-RAN with optimal fronthaul design," in *IEEE International Conference on Communications (ICC)*. Paris, France: IEEE, 2017, pp. 1–7.

[55] H. Yang *et al.*, "Multipath protection for data center services in openflow-based software defined elastic optical networks," *Optical Fiber Technology*, vol. 23, pp. 108–115, 2015.

[56] A. Nag, D. B. Payne, and M. Ruffini, "N: 1 protection design for minimizing olts in resilient dual-homed long-reach passive optical network," *IEEE/OSA Journal of Optical Communications and Networking*, vol. 8, no. 2, pp. 93–99, 2016.

[57] B. Chen *et al.*, "Cost-effective survivable virtual optical network mapping in flexible bandwidth optical networks," *Journal of Lightwave Technology*, vol. 34, no. 10, pp. 2398–2412, 2016.

[58] A. Tomassilli, B. Jaumard, and F. Giroire, "Path protection in optical flexible networks with distance-adaptive modulation formats," in *22nd International Conference on Optical Network Design and Modeling (ONDM)*. Dublin, Ireland: IEEE, 2018, pp. 30–35.

[59] T. Taleb, A. Ksentini, and B. Sericola, "On service resilience in cloud-native 5G mobile systems," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 3, pp. 483–496, 2016.

[60] C. Colman-Meixner *et al.*, "Resilient cloud network mapping with virtualized BBU placement for cloud-RAN," in *International Conference on Advanced Networks and Telecommunications Systems (ANTS)*. Bangalore, India: IEEE, 2016, pp. 1–3.

[61] B. M. Khorsandi *et al.*, "Survivable BBU Hotel placement in a C-RAN with an Optical WDM Transport," in *13th International Conference of Design of Reliable Communication Networks (DRCN)*. Munich, Germany: VDE, 2017, pp. 1–6.

[62] B. M. Khorsandi, F. Tonini, and C. Raffaelli, "Design Methodologies and Algorithms for Survivable C-RAN," in *22nd International Conference on Optical Network Design and Modeling (ONDM)*. Dublin, Ireland: IEEE, 2018, pp. 106–111.

[63] M. Y. Lyazidi *et al.*, "A Novel Optimization Framework for C-RAN BBU Selection based on Resiliency and Price," in *86th Vehicular Technology Conference (VTC-Fall)*. Toronto, Canada: IEEE, 2017, pp. 1–5.