Analyzing the quality of service of quantum oriented open radio access networks in 5G and 6G

By

Yassir Ameen Ahmed Al-Karawi

A Thesis Submitted in Partial Fulfilment of the

Requirements for the Degree of

Doctor of Philosophy

Department of Electronic and Electrical Engineering

College of Engineering, Design and Physical Sciences

Brunel University London

June 2024

To my beloved family and friends.

ACKNOWLEDGMENTS

I would like to express my gratitude and appreciation to the Ministry of Higher Education and Scientific Research (MOHESR), Cultural Attache, and University of Diyala in Iraq for their generous financing and invaluable cooperation in facilitating this project.

I would also like to express my profound gratitude to Professor Hamed S. Al-Raweshidy for his valuable advice, guidance, and support. The unwavering conviction, fervent endorsement, and constructive dialogues with him are really valuable and advantageous. The most highly valued aspect is his capacity for comprehension and patience, as he affords me ample autonomy in selecting my own areas of study.

Furthermore, I highly appreciate the significant role that my family has played in providing support, love, inspiration, and encouragement during my academic journey.

I would want to express my profound gratitude towards my friends and colleagues who have consistently provided me with encouragement during the duration of this project.

TABLE OF CONTENTS

Page

Page

LIST OF TABLES

LIST OF FIGURES

Figure Page

SYMBOLS

P_{CU}	central unit PC
$P_{CU}^{initial}$	central unit initial PC
P_{CU}^{T}	central unit total PC
P_{HV}	hypervisor PC
P_{dc}^{cu}	central unit DC PC
P_{ac}^{cu}	central unit AC PC
P_{cool}^{cu}	central unit cooling PC
P_{RF}	radio unit PC
P_{amp}	amplifier unit PC
P_{ORAN}	ORAN PC
D_{s}^{p}	processing delay for sending RRH
D_{m}^{p}	processing delay for MME
D_{mt}	delay of target MME
D_{qd}	delay of quantum network
$P_{traditional}$	traditional power consumption
$P_{quantum}$	quantum power consumption
P_{QT}	total PC
P_{server}	server PC
$P_{server}^{initial}$	initial PC of server
P_{RRH}	RRH power consumption
σ_{AC}	loss factor of AC

σ_{DC}	loss factor of DC
$\sigma_{cooling}$	loss factor of cooling
P_{AMP}	amplifier power consumption
$CRate$	data rate of traditional network
$QRate$	data rate of quantum network
$EE_{traditional}$	traditional energy efficiency
$EE_{quantum}$	quantum energy efficiency
$\|\psi\rangle$	quantum bit
P_{rb}^{cu}	PC of resource block at the CU
P_{rb}^{du}	PC of resource block at the DU
P_{msg}^{cu}	PC of message at the CU
P_{msg}^{du}	PC of message at the DU
RB_{msg}^{cu}	number of RB message at the CU
RB_{msg}^{du}	number of RB message at the DU
$Bits_{rb}^{du}$	number of bits in the DU
P_{supply}^{cu}	PC at the CU
$PC_{quantum}$	PC of quantum network
PC_{Laser}	PC of laser
PC_{Det}	PC of detector
PC_{Driver}	PC of driver
$\tau_{traditional}^{cu}$	delay of traditional CU
$\tau_{quantum}$	delay of quantum CU
C_q	Capacity of quantum network CU
L_{rb}^{x}	number of bits in each RB at modulation index x
P_{vm}	PC of the VM
P_{VM}	PC of total VMs
P_{ent}	PC of entanglement network

ABBREVIATIONS

BBU	base band unit
BS	base station
CAPEX	capital expenditure
CA	carrier aggregation
CN	core network
CPU	central processing unit
CPRI	common public radio interface
C-RAN	cloud radio access network
CU	central unit
CoMP	coordinated multi point
COTS	commercial, off-the-shelf
EE	energy efficiency
eMBB	enhanced mobile broadband
eNodeB	evolved NodeB
EPC	evolved packet core
EPS	evolved packet system
E-UTRAN	evolved universal telecommunication radio access network
FDM	frequency domain multiplexing
FFT	fast fourier transform
GOPS	giga operation per second
GSM	global system for mobile communications

HDD	hard disk Drive
HetNet	heterogeneous network
H-CRAN	heterogeneous CRAN
HSPA	high speed packet access
HSS	home subscriber server
HV	hypervisor
LTE-A	LTE Advanced
LTE	long term evolution
MIMO	multiple-input multiple-output
MME	mobility management entity
mMTC	massive machine-type communication
MCS	modulation and coding scheme
mmWave	millimeter wave
NAS	network attached storage
NFV	network function virtualisation
NIC	network interface controller
ORAN	open radio access networks
OFDMA	orthogonal frequency division multiple access
OPEX	operational expenditure
PC	power consumption
PCRF	policy and charging rules function
PGW	packet gateway
PM	power model
QKD	quantum key distribution
QoS	quality of service
Qubit	quantum bit
RAM	random access memory
RF	radio frequency
RRC	radio resource control

RRH	remote radio head
Rof	radio over fiber
RU	radio unit
SAE	system architecture evolution
SC-FDM	single carrier-frequency domain multiplexing
SDN	software defined network
SE	spectral efficiency
SGW	serving gateway
SNR	signal-to-noise ratio
SotA	state of the art
SP	service provider
SQP	sequential quandratic programming
UE	user equipment
URLLC	ultra-reliable low-latency communications
UMTS	universal mobile telecommunication system
VM	virtual machine
VoLTE	voice over long-term evolution
X2-AP	X2-application protocol
3GPP	third mobile generation partnership project
6G	sixth generation
5G	fifth generation
4G	fourth generation

Abstract

The Open Radio Access Network (ORAN) technology has been developed to provide efficient spectrum sharing and cost-effective solutions, but it also increases network infrastructure and power usage. To evaluate the performance of the power consumption (PC), a model was proposed to quantify the compromises associated with virtualizing a server within an ORAN infrastructure, considering factors like the quantity of virtual machines, allocation of system resource blocks, and bandwidth. However, virtualization of the network has resulted about 50% reduction in the total PC in comparison with traditional cloud networks. However, the ORAN paradigm has produced more PC compared to the virtualization case, about 30% in the total PC and 10% in the cooling PC. Unless the advantages of ORAN are fully realized, the addition of extra units within the ORAN, specifically the DU servers, may result in more PC that might advocate against the ORAN.

Subsequently, a work was proposed to examine the criteria and evaluations for prospective quantum solutions in conventional ORAN networks, focusing on entanglement phenomena to enhance the efficiency of the X2 application (X2-AP) protocol. This approach reduces the overhead of X2-AP signaling, reducing time and power consumption associated with standard cloud-based systems. As a result, increasing the number of photons has decreased the delay to about 40% compared to the traditional ORAN network. In addition, the energy efficiency in the quantum case has been increased while decreasing the power consumption by about 10%.

This study also investigates the use of quantum entanglement-based approaches and their influence on conventional ORAN architecture's signaling among the central units (CUs) and distributed units (DUs) by replacing the traditional method with entangled photons. The results showed that the proposed method has promised about 45%, 40% and 10% reductions in the EE, PC and delay, respectively, when compared to the traditional ORAN.

Finally, a novel methodology for addressing problems in ORAN and implementing load balancing algorithms is presented, focusing on selecting ORAN servers with lower energy efficiency for quantum load balancing. This comparative analysis offers valuable insights for developing power-efficient ORAN implementations. This nonlinear problem was solved using Lagrange multiplier method to solve the problem mathematically, and using the Matlab (fmincon) software to solve the problem numerically. In which, two algorithms are used, sequential quadratic programming (SQP) and active-set. It was shown that the SQP model exhibits superior energy efficiency compared to the active-set model, with a difference of approximately 45%.

Chapter 1: Introduction

1.1 Overview

As we go through the fast developments in telecommunications, the increasing number of networked devices continues to rise, fueled by the growing need for applications and services that need a large amount of bandwidth [1]. By 2024, it is projected that the global number of internet-connected devices will surpass 50 billion [2]. The extensive network of devices drives a rapid growth in mobile data traffic, posing intricate issues that need inventive solutions in network architecture and administration [3, 4].

The introduction of 5G technology has brought about a crucial shift in dealing with these difficulties, offering far improved capabilities compared to its predecessors. 5G networks are designed to support a wide array of applications, from enhanced mobile broadband (eMBB) to ultra-reliable low-latency communications (URLLC), and massive machine-type communications (mMTC). These technologies enable innovative applications in industries like as autonomous driving, industrial IoT, and urban infrastructure management, establishing a higher level of connection and dependability [5, 6]. The telecoms sector is currently preparing for the development of 6G, in anticipation of future developments. Anticipated to launch approximately in 2030, 6G seeks to incorporate nascent technologies like artificial intelligence and quantum computing into the network infrastructure. This integration has the potential to significantly improve network security, data transfer speed, and user connection, enabling a broader array of applications and services. In order to meet this demand, additional base stations (BS), including Macro, Pico, and Femto, are deployed, leading to an increase in power consumption (PC).

The Long Term Evolution (LTE), commonly known as 3GPP Release 17, is the most recent iteration of the Third Generation Partnership Project (3GPP). The objective of this version is to address the limitations observed in previous generations of mobile communication technologies, like GSM, UMTS, HSPA, and others. These systems engage in competition for a minimum duration of 10 years prior to the emergence of LTE. The primary distinctions among these systems lie in the utilisation of IP packet design by LTE and the introduction of a novel air interface known as orthogonal frequency division multiple access (OFDMA). Moreover, it establishes novel quality of service (QoS) bearers in order to ensure the fulfillment of user equipment (UE) criteria. The Long-Term Evolution (LTE) system has been introduced with the objective of establishing a competitive presence in the market for a minimum duration of ten years. The primary goals of this system were:

- Less latency.
- Network architecture for communication that is packet-oriented.
- Adjustable frequency ranges of 1.25 to 20 MHz.
- Higher data speeds, including transmission at up to 100 Mbps downlink and 50 Mbps uplink.
- Using single carrier-frequency division multiple access (SC-FDMA) for uplink transmission and OFDMA multiple access for downlink.

1.2 Long Term Evolution paradigm

The evolved universal telecommunication radio access network (E-UTRAN) has emerged as a result of the evolution of the UMTS legacy system in LTE. The evolved packet system (EPS), which is made up of this and the system architecture evolution

(SAE) that includes the evolved packet core (EPC), is referred to as the LTE network architecture. The segmentation of data into smaller units known as packets is a component of the packet system. The designated host will then receive these packets after they have been transferred over the network. Routing algorithms make it possible for each packet to take a unique path through the network during transmission in order to get to its intended location. This contrasts with circuit switching systems, where the path is established and decided upon before the connection is made. According to Figure 1.1, the UEs, E-UTRAN, and EPC make up the core three elements of the LTE architecture [7]. The UEs and the network are connected through a number of eNodeBs, which make up the E-UTRAN network. These eNodeBs are in charge of controlling the resource allocation needed for communication as well as processing the data received from the UEs. The S1-U interface enables the user equipment's data plane to be routed from the eNodeB to the serving gateway (SGW) and then back to the eNodeB where the other user equipment is located once the call has been initiated. The S1-MME control plane interface is used by the eNodeB to update the mobility management entity (MME), in contrast. The SGW relies on the MME to update its database, facilitate critical mobility functions, and provide updates regarding the location of the UE to the SGW. UE packets are forwarded and routed to their intended destinations largely by the SGW. The SGW enables the transport of packets to the Packet Gateway (PGW) in the event that the UE demands connection to additional networks or the Internet. PGW is in charge of a number of duties, such as inter-operator pricing, packet filtering, and IP distribution [8].

The second component comprises a multitude of evolved NodeBs that are interconnected and communicate with one other over the X2-interface to facilitate essential handover procedures. Additionally, the S1 interface is utilised to establish connections between these eNodeBs and the EPC. The eNodeB serves as an intermediary to establish a connection between the UE and the core network, while also offering the essential protocols. The eNodeB assumes the responsibility of resource allocation, namely in terms of time and frequency dimensions, for multiple UEs. This allocation is carried out while maintaining the necessary QoS standards for the UEs. Furthermore, the eNodeB possesses mobility management capabilities, including handover signaling and radio link metrics [8].

1.2.1 Core Network

The core network, alternatively referred evolved packet core, EPC, comprises three primary entities: the MME, the SGW, and the PGW. In addition, certain logical entities, such as the policy and charging rules function (PCRF) and the home subscriber server (HSS), may be regarded as components of the CN. Each individual entity is accountable for distinct functions. The MME assumes the responsibility of facilitating inter CN mobility signalling with other 3GPP networks. Additionally, it oversees tasks like as authentication, authorization, time zone signalling, and bearer management. The latter include operations related to the activation and deactivation of bearers. An EPS carrier refers to the manner in which UE communication is handled as it traverses the network [9].

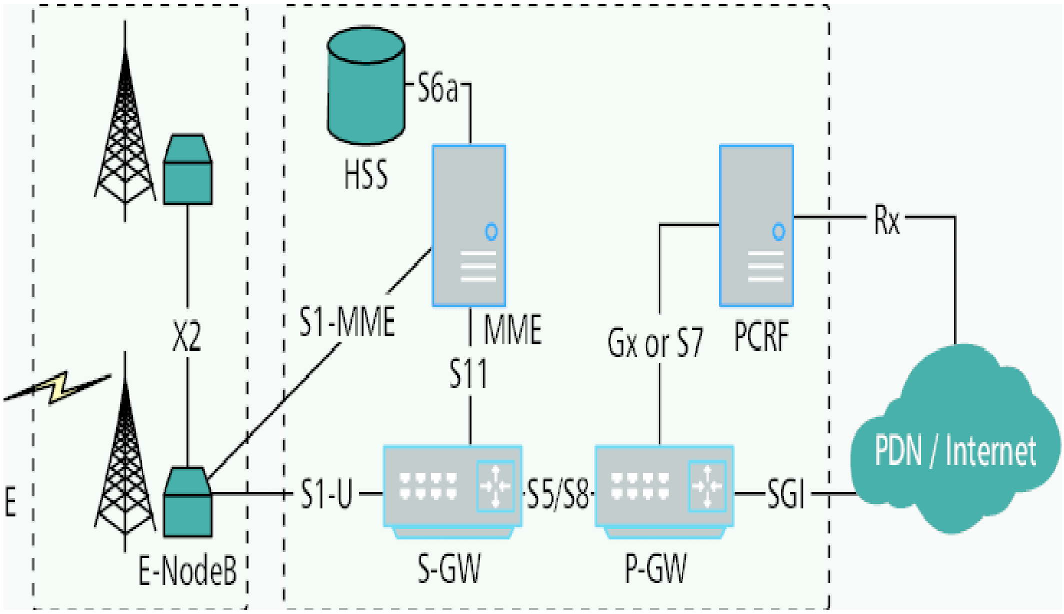

Fig. 1.1. The network architecture of LTE [10].

1.3 Long Term Evolution Advanced

To improve the functionality of the Long-Term Evolution (LTE) system, a number of different techniques have been developed, which has led to the creation of an improved iteration known as LTE-Advanced (LTE-A), which is also referred to as 3rd Generation Partnership Project (3GPP) Release 10. These techniques were utilised by LTE in order to raise data transfer rates, broaden coverage, improve throughput, and decrease latency, all of which contributed to an increase in user satisfaction (LTE Advanced: Next Generation Mobile Broadband, n.d.). In addition to this, LTE-A is able to offer support for heterogeneous networks, in which the layout of the Macro cell integrates low power nodes such as Pico cells, Femto cells, and relays. On the other hand, these technologies fall under the following categories:

1.3.1 Voice over Long-Term Evolution

Voice over Long-Term Evolution (VoLTE) is a technology that enables the transmission of voice calls across the 4G LTE network, offering superior voice quality and faster call setup times compared to traditional circuit-switched voice communications. VoLTE provides high-definition audio, less background noise, and enhanced audio clarity, making it an ideal choice for internet browsing, application usage, and data exchange during telephonic conversations. VoLTE also offers expedited call setup duration, enhancing responsiveness and efficiency during the calling process. It allows simultaneous voice and data transmission, enabling users to use data services while engaged in a call, which is particularly beneficial for internet browsing, application usage, and data exchange. Additionally, VoLTE has been found to exhibit greater efficiency compared to conventional voice services, potentially prolonging mobile device battery life [11]. Seamless handover between LTE cells and Wi-Fi networks (VoWiFi) is another advantage of VoLTE. Wideband audio codecs like Adaptive Multi-Rate Wideband (AMR-WB) are often used to expand the audio frequency spectrum and enhance conversational authenticity. VoLTE also offers expanded capabilities, including video calling and real-time multimedia content sharing. Compatibility with previous 2G and 3G networks ensures effective voice communication even in regions with limited LTE coverage. VoLTE calls include encryption, providing a higher level of security compared to conventional calls.

The Quality of Service (QoS) feature in VoLTE networks prioritizes voice traffic, ensuring consistent call quality even in high data usage situations. VoLTE has gained widespread prevalence in numerous nations due to the maturation of LTE networks. To use VoLTE technology, mobile devices must have necessary compatibility, be extended by their service provider, and enable VoLTE functionality within their settings. However, availability may vary depending on geographical region and mobile network provider [12].

1.3.2 Advanced multiple-input multiple-output

The term "Multiple-input multiple-output (MIMO)" refers to a type of communication system in which both the transmitter and receiver are equipped with multiple antennas. The utilisation of MIMO technology played a pivotal role in the triumph of LTE and several competing systems. In the case of LTE, the MIMO system employed was capable of accommodating up to four layers. Moving on, the progression involves Long-Term Evolution Advanced (LTE-A). Based on the specifications outlined in LTE-A standards, the attainment of maximum spectrum efficiency necessitates the utilisation of 8×8 spatial multiplexing. In contrast to LTE, LTE-A has uplink MIMO technology, which allows for the utilisation of up to four layers for a single UE [13].

1.3.3 Carrier aggregation

The objective of the suggested approach is to achieve a peak rate of around 1 Gbps by combining various frequency bands to produce a wider bandwidth. This bandwidth is designed to serve the needs of a single piece of UE. For example, if a SP holds a bandwidth allocation of 10 MHz in the 800 MHz spectrum and an additional 20 MHz in the 1900 MHz spectrum, it is possible for these two bands to be combined to generate a consolidated bandwidth of 30 MHz. This would be the case in a scenario where the SP possesses both of these spectrums. The numerical value has the ability to increase until it hits a maximum threshold of 100 MHz, at which point it will be considered to have reached its maximum. There are two main sorts of CA strategies, as described in carrier aggregation for LTE reference. The first is continuous CA, in which the carrier components are positioned adjacently. The second type is called non-continuous CA, and it refers to situations in which the multiple carrier components are separated by gaps across the bandwidth [14]. The first type, in general, is the one that can be carried out in a manner that is somewhat simpler. Backward compatibility with the LTE system does not need the texture of the LTE physical layer to be altered in any way. This is because there is no need for such an alteration. Utilising a single module for the fast Fourier transform (FFT) and a single radio frequency (RF) component is all that is required to accomplish this goal. In addition to this, this has an effect on how easy it is to manage algorithms and allocate resources in the first form of data, in comparison to the second kind of data. However, due to the restricted resources that are available, the practical implementation of continuous cellular automata is made more difficult in settings with low bandwidth. Therefore, the second category is the one that shows to be more realistic as it allows operators to successfully use their existing spectrum resources. These resources include both fragmented and unutilized frequency bands, in addition to those bands that are now assigned for outdated systems [15].

1.3.4 Coordinated multi point

Coordinated multi point (CoMP), as defined in the 3rd Generation Partnership Project (3GPP) Release 11, is a significant technology in Long-Term Evolution-

Advanced (LTE-A) that enhances the average capacity of cellular networks, enhances the signal-to-noise ratio (SNR), and optimises the utilisation of system resources and spectral efficiency. The Coordinated Multi-Point (CoMP) technique facilitates communication between numerous eNodeBs by dynamically mitigating the interference caused by transmission signals. As a result, the efficiency of the downlink is significantly improved when the signals transmitted by several eNodeBs work together to mitigate the interference between cells [16]. The utilisation of the same frequency spectrum across all BS sectors results in heightened interference experienced by user equipment located at the periphery of a cell, particularly when receiving signals from many BSs simultaneously. The utilisation of CoMP enables the collaboration of sectors within a single BS through intra-site coordination, while inter-site coordination occurs among numerous or neighbouring eNodeBs. The application of COMP is observed in both the downlink and uplink. However, the deployment of All (Coordinated Multipoint) technology brings up a number of challenges in terms of backhaul demand. These challenges encompass various aspects such as low latency, high cell capacity, increased synchronisation, higher complexity, as well as increased channel estimate and PC [17].

1.3.5 Heterogeneous networks

Relay nodes are strategically positioned at the periphery of cellular networks in order to enhance network capacity and improve coverage. The function of these low power stations is to serve as repeaters, with the primary objective of rebroadcasting signals that have been received or transmitted, so enhancing the quality of the signal [18]. The relays establish connections with the eNodeBs over wireless networks. The presence of such entities provides significant cost savings in comparison to the implementation of new eNodeBs. This concept is closely related to the notion of a heterogeneous network (HetNet). In order to enhance the capacity and coverage of networks, HetNets facilitate the integration of cells with diverse sizes, each equipped with distinct radio access technology and output power, to operate in cooperation. The deployment of small-sized cells within the coverage area of large-sized cells allows for the simultaneous servicing of a greater number of UE and enhances the quality of SP to UEs located at the periphery of the cell.

In addition to the aforementioned technologies, several other evolutionary proposals have been offered for LTE-A. These include self-organizing networks, cognitive radio, and enhanced inter-cell interference coordination [19].

1.4 Fifth Generation

Mobile operators are consistently in search of novel concepts, architectures, protocols, and advanced digital signal processing (DSP) methods to efficiently manage the rapid increase in data demand. This is done while also ensuring scalable and expedient connectivity [20].

According to [21], by the year 2030, the 6G wireless and mobile communication systems would be capable of delivering approximately 1000 times greater capacity than 5G networks. This advancement would be accompanied by a reduction of up to 90% in energy consumption. Furthermore, less than 1ms of end-to-end latency will be achieved. Furthermore, it was anticipated that the battery life of connected devices would increase by a factor of 10, and the end-to-end latency would be reduced by a factor of 5 when compared to the existing fourth generation (4G) system. Nevertheless, the majority of futuristic algorithms commonly depend on the practise of over-provisioning resources in order to guarantee the fulfilment of specified needs. This includes techniques like spatial densification and spectral aggregation [22]. Consequently, this results in elevated PC and thus, increased expenses for the network providers. In order to surpass these expenses, it becomes imperative to use novel paradigms and network designs [23]. The aim of these new technologies is diverse. In this context, a concise description of some essential technologies for 5G is provided [24].

1.5 Cloud radio access networks

The C-RAN design has been proposed by various operators, including NTT, KT, and France Telecom/Orange. The companies involved in the telecommunications industry include Telefonica, SoftBank/Sprint, and China Mobile, together with equipment vendors such as Alcatel-Lucent, Light Radio, Nokia-Siemens, and Liquid Radio.

C-RAN is recognised as a fundamental technology that is essential for the implementation of high-performance 5G networks [25]. The C-RAN architecture represents an enhanced iteration of conventional network paradigms, incorporating the principles of cloud computing into mobile systems [26], [27]. C-RAN is a network architecture that comprises several remote radio heads (RRHs) and a baseband unit (BBU) pool, as shown in Fig 1.2. The RRHs, which operate at low power, are spread and interconnected to the BBU pool either over high-bandwidth optical fibres or wireless links. The latter, in this case, serves as the primary location where numerous BBU servers are hosted, including those in the cloud and data centre. The adoption of this technology ensures a decrease in both operational (OPEX) and capital (CAPEX) expenses. This is achieved through the reduction of site visits, maintenance, and leases, resulting in a perpetual decrease in the overall cost of operating the network [28]. In addition, this technology facilitates dynamic allocation of resources between BBUs and RRHs, so enabling off-loading algorithms to distribute processing load among adjacent BBUs. Consequently, certain BBUs can be deactivated in order to conserve energy. The cooperative efforts within the aquatic environment contribute to the improvement of the system's overall data transmission capacity, as well as its ability to efficiently utilise the available frequency spectrum and energy resources [29]. Additionally, the cooling demand can be minimised in the context of C-RAN due to the presence of a limited number of distributed pools spanning a significantly vast geographical region. Within each pool, there will exist a limited number of cooling units to accommodate a large number of BBUs. at contrast, at typical BS sites, each BBU need an own cooling unit.

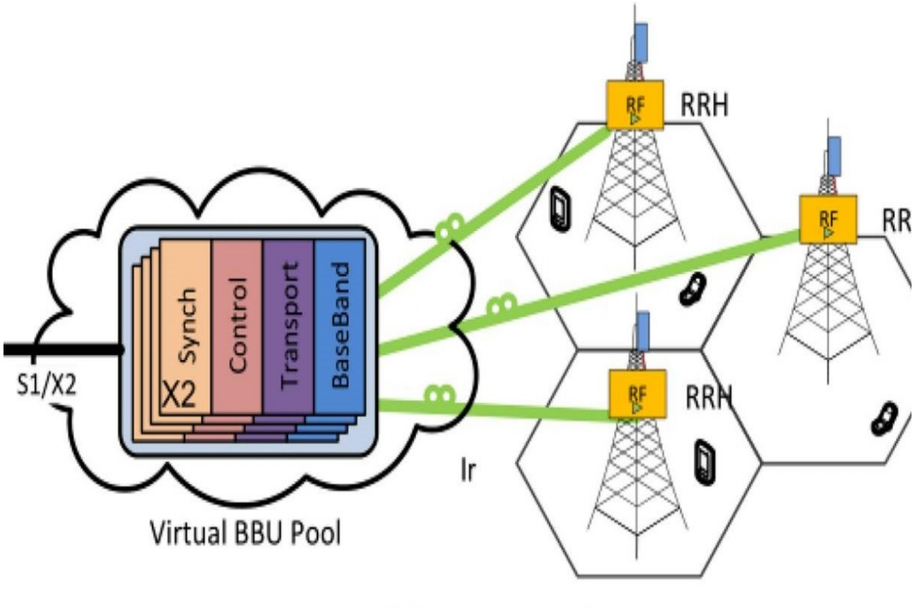

Fig. 1.2. The CRAN architecture [30].

1.6 Open RAN

The introduction of ORAN has led to a significant transformation in the field of wireless communication infrastructure. Traditional cellular networks often exhibit distinct features of closed, integrated structures, where many components obtained from a single supplier are closely interconnected, hence restricting flexibility and innovation. However, the ORAN design presents a new method that involves separating and organising different components inside a radio access network [31].

Open RAN is primarily based on the concepts of disaggregation and standardisation, which enable the separation and independent scalability of the core component of the network. This architecture strategically segments the network into three primary elements:

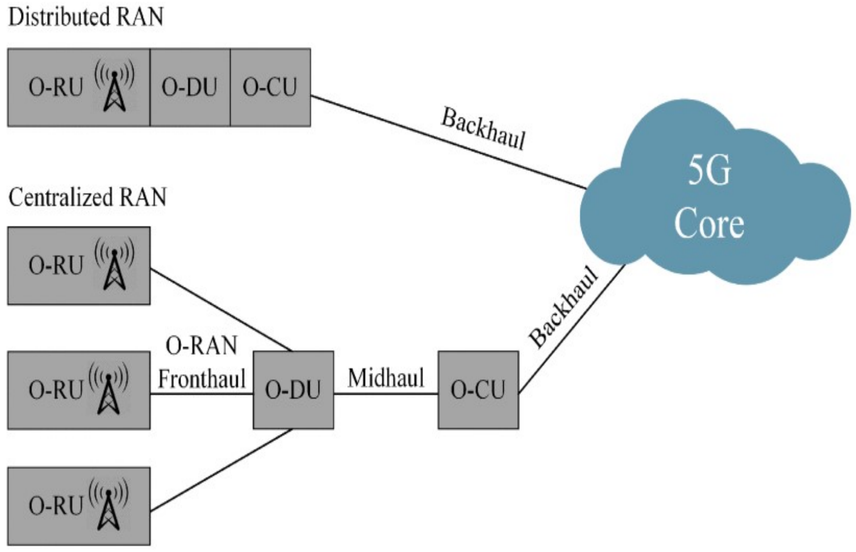

Fig. 1.3. The Open RAN architecture [32].

1. **Radio Units (RUs)**: These units are responsible for managing the radio frequency signals and are essential for the transmission and reception of data across the network. In an Open RAN configuration, Remote Units (RUs) are intentionally built to be agnostic to any one manufacturer, enabling them to function smoothly with hardware provided by several providers. Interoperability is essential for increasing the adaptability of systems and expanding the range of technology implementation.

2. **Distributed Units (DUs)**: DUs play a crucial role in the baseband processing layer by overseeing the real-time execution of network activities. They have a strong integration with the RUs but may be easily installed at other geographical locations. The distribution is facilitated by open interfaces that allow Dynamic Units (DUs) to seamlessly interact with Remote Units (RUs) produced by various manufacturers, hence enhancing the operational efficiency and reaction times of the network.

3. **Centralized Units (CUs)**: CU's serve as the central processing unit of the network, responsible for managing the control plane operations of the Radio Access Network (RAN). They provide communication and synchronisation among many DUs, consolidating the control signals and guaranteeing seamless network operation. By separating CUs from DUs, it becomes possible to centrally control and scale network resources. This is crucial for effectively managing large-scale network operations and meeting the demands of modern applications.

The architectural design prioritises openness and flexibility by including unique components, enabling network operators to seamlessly connect and use equipment from different manufacturers. ORAN facilitates a competitive atmosphere by using open interfaces, which leads to reduced costs and faster implementation of cutting-edge technology. This technique not only improves the adaptability and expandability of network installations but also enables the quick incorporation of new technologies into the RAN [32–34].

Cloud RAN aims to centralise the baseband processing in a centralised pool, or cloud, to achieve benefits in terms of operating costs and efficiency. ORAN, however, incorporates open interfaces and virtualization (VM), enabling more flexibility and interoperability across equipment from various suppliers. ORAN open interfaces allow for more flexible and cost-effective installations compared to Cloud RAN, which may use proprietary interfaces.

The ORAN promotes interoperability and fosters the participation of various vendors through the adoption of open interfaces and adherence to defined standards. Moreover, this approach aligns with the broader movement towards the deployment of software-defined and virtualised networks. The employment of software modules on commercially accessible hardware offers a potential avenue for conducting network operations, leading to reduced deployment costs and enhanced scalability [35, 36].

However, akin to every groundbreaking concept, the ORAN faces numerous hurdles. This thesis addresses and examines several of the aforementioned difficulties, providing analysis and proposing solutions.

1.7 Motivation of the thesis

The telecommunications sector is being transformed by the emergence of 5G and subsequent generations of wireless communication technologies, which provide significant advancements in terms of data speeds, latency, and device connectivity. The Open Radio Access Network (ORAN) paradigm is important to this transformation, as it provides a software-centric, disaggregated, and open method to constructing wireless networks. This thesis seeks to examine the significant importance of QoS in the context of 5G, 6G and subsequent generations of wireless communication technology. It also aims to explore the intricate nature and diverse characteristics of ORAN architectures, as well as the potential for enhancing performance optimisation. QoS assumes a pivotal part in the effective implementation and functioning of these networks. Consequently, comprehending and enhancing QoS inside ORAN designs is important for their triumphant deployment and operation. The ORAN strategy promotes the establishment of a vendor-neutral interoperability framework, which facilitates the deployment of several vendors' equipment. Gaining a comprehensive understanding of QoS in the context of ORAN holds the potential to foster the advancement of novel solutions aimed at augmenting network performance, reliability, and efficiency.

1.8 Aim of the thesis

One of the primary concerns within the realm of 5G paradigms is the optimisation of several factors, including enhancing EE, optimising radio planning, minimising latency, and lowering processing complexity. The costs associated with electricity bills for SPs and call charges for customers exhibit a linear relationship with respect to the quantity of data delivered and the specific type of service utilised. SPs are motivated to enhance the QoS for UEs through the development of novel technologies and algorithms, as well as the expansion of cell deployment. Nevertheless, the act of upgrading their PC contributes to an escalation in their monthly expenses, thus leading to detrimental consequences for the environment, including the exacerbation of global warming. This also pertains to the implementation of increased client costs as a means of compensating for such losses. Hence, it is imperative to direct attention on the PC in forthcoming hybrid and heterogeneous networks. According to a study by [29], the BS accounts for 80% of the energy consumption in network operations. In order to address this issue, future power reduction strategies are being explored, such as C-RAN, SDN-based C-RAN, virtualised networks and quantum mechanics based methods. These technologies fit into the broader network architecture by enabling more efficient resource management and reducing the overall energy footprint. In the context of these networks, it is of utmost importance to prioritise the reduction of PC and latency through the use of cooperative BBUs, PC reduction methods, and offloading mechanisms. Nevertheless, the validity of these strategies remains uncertain unless their practicality and associated compromises are thoroughly understood.

1.9 Challenges and Objectives

The primary objective of research endeavours is to advance and innovate upon existing state of the art (SotA) methodologies, hence enhancing network performance. Enhancing the EE and delay is seen as a significant component that influences the overall performance. This involves a dual approach of increasing the bit rate of the system while simultaneously reducing the PC. In recent years, there has been significant interest in developing a dependable approach for measuring the PC of networks, as opposed to introducing novel algorithms aimed at reducing PC. The subsequent items succinctly illustrate the significance of doing such an evaluation:

1. If there is a proposal to include new components into the SotA network design, such an addition has the potential to improve the network's flexibility, programmability, or system bit rate. Simultaneously, it could potentially contribute to a significant increase in energy usage. Hence, it is imperative to do an EE evaluation in order to ascertain the cost and assess the PC.

2. The implementation of advanced strategies that enhance coordination and cooperation within the network includes dynamic allocation of network resources in response to traffic demand, enabling the integration of new services, distributing the workload evenly across servers within the cloud, and implementing sleep mode for certain units within the network. The implementation of these technologies necessitates the utilisation of a dedicated device or server, which subsequently introduces a significant energy overhead to the system and diminishes its EE. The potential for unforeseen energy expenses to surpass the anticipated benefits of the proposed technology necessitates the evaluation of these costs prior to its adoption.

3. In order to improve the EE of SotA systems, it is advisable to provide a comprehensive platform for evaluating the performance of these systems on PC. This platform would facilitate the comparison of PC measurements between different proposed systems, thereby enabling a clear understanding of the power gain achieved. However, the integration of quantum mechanics into communication systems represents a significant advancement. Therefore, it is imperative to assess the enhancements it brings in terms of latency and PC.

4. Certain improvements have the potential to improve EE but may have a negative impact on network performance. Nevertheless, by a thorough assessment of the PC, one may determine whether it is fair to sacrifice network performance in exchange for a decrease in PC.

5. Gaining a comprehensive understanding of the PC patterns exhibited by VMs might facilitate the development of novel approaches aimed at optimising resource allocation and enhancing the hypervisor (HV) scheduling mechanisms employed among these VMs. In the context of ORAN, it is important to note that the VMs are not exclusively owned by a single SP. Instead, each VM is intended to cater to several providers. It is crucial to comprehend the power and delay aspects of such design.

SPs tend to avoid allocating excessive time and resources towards assessing network PC and delay, as it can be properly inferred from their monthly electricity payments and witnessing lack of excellent service. Furthermore, they allocate supplementary resources towards the advancement of innovative algorithms and optimisation methodologies aimed at reducing PC and efficiently controlling the energy usage of their appliances and network, as well as providing high speed communications. It is recommended that these tactics be adopted in any situation when there is a rise in network or device usage. Nevertheless, it is imperative to assess the power and delay attributes of components and devices driven by a specific requirement, as this information is necessary for identifying areas of innovation. In this context, the effectiveness of models for power and delay assume paramount importance.

Not to forget, in order to tackle the computational complexity associated with power and delay calculations in the field of literature, it is necessary to develop models that are more brief, precise, and parameterized.

To exceed these challenges, the following objectives have been concentrated on:

1. Offering a model for assessing the PC of the ORAN paradigm is of utmost importance. This model plays a critical role in calculating the potential decrease or increase in PC compared to networks, as well as in evaluating future tactics aimed at reducing PC.

2. Developing a model for quantifying the PC associated with the implementation of quantum devices in the ORAN paradigm. This modelling approach is essential for assessing the extent to which the integration of quantum devices laser and detector) affects the overall PC in ORAN. Moreover, it provides a comprehensive framework for assessing the emerging field of quantum computing and its novel algorithms in a cloud-based environment.

3. The utilisation of the entanglement phenomenon is proposed as a means to enhance the performance of the X2 application (X2-AP) protocol, which is a communication protocol used between eNodeBs in LTE networks to manage

functions such as load balancing, handover management, and inter-cell interference coordination, enhancing overall network efficiency and connectivity. This enhancement is achieved by reducing overhead signalling, which is typically associated with time and energy consumption in traditional cloud systems. Our objective was to minimise delays by incorporating quantum techniques into ORAN, achieved through the analysis and comparison of latency in both paradigms.

4. The addition of new devices, protocols, standards, and modifications to the legacy network by ORAN leads to an increase in PC and latency costs for both distributed units and central units. The study presented an approach based on quantum entanglement to demonstrate its influence on the signaling of the standard ORAN architecture. Quantum entanglement is a phenomenon where particles become interconnected so that the state of one particle instantaneously affects the state of another, regardless of the distance between them.

1.10 Thesis Contributions

1. In Chapter 3, the ORAN technology has undergone significant development as a means to provide efficient spectrum sharing and cost-effective solutions. Nevertheless, this system will lead to an increase in both network infrastructure and power usage. In order to evaluate the performance of a PC, it is necessary to employ a model that quantifies the compromises associated with virtualizing a server within an ORAN infrastructure. Various criteria, including the quantity of VMs, the allocation of system resource blocks (RBs), and the bandwidth, have been employed and contrasted with both bare virtualization and conventional cloud networks.

2. In Chapter 4, this work examines the essential criteria and evaluations required for prospective quantum solutions in the context of conventional ORAN networks. Specifically, the utilisation of entanglement phenomena is proposed as a

means to enhance the efficiency of the X2 application (X2-AP) protocol. This enhancement is achieved by reducing the overhead of X2-AP signalling, which encompasses the time and energy consumption often associated with standard cloud-based systems. Our objective was to minimise delays by incorporating quantum techniques into cloud computing, achieved through the analysis and comparison of latency in both paradigms.

3. In Chapter 5, the implementation of ORAN introduces more devices, protocols, and standards, resulting in a transformation of the existing design. This transformation, however, leads to an increase in PC and latency costs. The expenses arise throughout the signalling procedure of several components within the ORAN framework, including the distributed unit (DU) and central unit (CU). This work aims to investigate the utilisation of a quantum mechanics-based approach, particularly entanglement theory, and examine its influence on the conventional ORAN architecture's signalling. Furthermore, this study presents a comprehensive analysis of the performance metrics, including processing capacity, latency, and EE, for both conventional and quantum-based ORAN systems. The comparison of these systems takes into account many network parameters, including the quantity of signalling messages and the given bandwidth.

4. In Chapter 6, a novel methodology was presented for addressing problems in ORAN and implementing load balancing algorithms. The optimisation challenge involves the selection of ORAN servers with lower EE for the purpose of quantum load balancing. This is achieved through the utilisation of the Lagrange multiplier approach and the numerical problem-solving technique known as 'fmincon'. This chapter conducts a comparative analysis between sequential quadratic programming and active-set approach methodologies, with the aim of offering valuable insights for the development of power-efficient ORAN implementations.

1.11 Thesis Organization

In Chapter 2, the available literature was presented. In Chapter 3, ORAN power model was presented. Chapter 4 depicts the quantum based X2-AP signalling overhead is analysed. Subsequently, in Chapter 5, CU-DU signalling procedure is analysed. Subsequently, Chapter 6 presents a quantum load balancing technique assisted optimisation algorithm is proposed. Moreover, in chapter 7, the thesis was finalized and the findings were concluded. In addition, the subsequent trends and potential areas of research are deliberated.

Chapter 2: Literature survey

The ORAN architecture represents a versatile, modular, and interoperable strategy for constructing wireless networks. The system comprises several essential elements, namely the radio unit (RU), distributed unit (DU), centralized unit (CU), open interfaces, network controller, orchestration Layer, management and orchestration framework, virtualization and cloud-native elements, service management and security elements [37]. The radio unit is responsible for the transmission and reception of radio signals. On the other hand, the DU is tasked with processing these signals, carrying out functions such as digital processing, beamforming, and modulation/demodulation. The CU is responsible for overseeing and managing many advanced operations within the network, such as network control, administration, and orchestration. Open interfaces play a crucial role in facilitating interoperability among diverse components within a system. These interfaces, such as fronthaul, midhaul, X2, O1, A1, and policy management, enable automatic resource allocation and optimisation [38]. The network controller, alternatively referred to as the RAN intelligent controller, is responsible for the implementation of network policies and the allocation of resources in an intelligent manner. It utilizes real-time network conditions to make dynamic choices. The orchestration layer is responsible for the management and automation of the provisioning of Open RAN resources. It interacts with the CU and other network management systems in order to carry out its functions [39]. The management and orchestration unit is responsible for managing the resources within the ORAN system. It performs essential functions such as resource allocation, scaling, and service chaining. ORAN facilitates the virtualization of network services, hence enabling the deployment of cloud-native systems. Service management and application programming interfaces (APIs) facilitate the integration of services and the interaction with higher-level applications. This allows for the provision of a diverse array of services and applications to end-users [40].

The incorporation of security measures within the ORAN design is of paramount importance, encompassing encryption, authentication, and the implementation of security policies. These measures serve to safeguard the network from potential vulnerabilities and threats. The ORAN design facilitates the establishment of vendor-neutral interoperability, hence augmenting adaptability, mitigating the risk of vendor lock-in, and fostering innovation within the RAN domain [41].

2.1 Power consumption

The power consumption of ORAN exhibits significant variability due to several factors, encompassing the network's architectural design, scale, technological preferences, and operating circumstances. The following are several factors to be taken into account in relation to power consumption within the context of ORAN.

1. The first aspect to consider is the size of the network. The power consumption of the ORAN network is strongly influenced by the quantity of base stations, remote radio units, distributed units, and central units. In general, networks that possess a greater number of network pieces tend to exhibit higher power consumption [42].

2. Architecture of radio access network: The implementation of ORAN facilitates the incorporation of flexible network typologies. Virtualized or cloud-native radio access network designs have the potential to exhibit greater energy efficiency in comparison to conventional radio access networks. This is primarily due to their ability to facilitate dynamic resource allocation and scaling [43].

3. Hardware Efficiency: The energy efficiency of the hardware components utilised in ORAN is of paramount importance. The implementation of power-efficient hardware has the potential to effectively mitigate and decrease the overall power consumption [44].

4. Power Management: The implementation of effective power management measures, such as the capability to transition network devices into low-power or sleep modes during periods of reduced traffic, has the potential to decrease the overall power consumption [45].

5. Spectrum efficiency refers to the ability of a wireless communication system to utilise the available frequency spectrum in an efficient manner. The optimisation of spectrum allocation can have a significant influence on power consumption. The utilisation of radio resources can be optimised by the implementation of dynamic spectrum sharing and interference management strategies [46].

6. Impact of Environmental Conditions: The cooling and power demands of network equipment can be influenced by several environmental elements, such as weather conditions and temperature. The exacerbation of environmental conditions may result in a corresponding rise in energy demand [47].

7. optimisation of Software: The implementation of software optimisations, such as the utilization of intelligent algorithms for the purpose of managing resources, can effectively mitigate power consumption by dynamically adapting power levels and configurations [48].

8. Load balancing refers to the process of distributing workloads across many computing resources in order to optimize resource utilization, improve performance, and ensure high availability of services. Efficient load balancing facilitates the equitable distribution of traffic and workloads among network parts, hence mitigating the risk of overload and minimizing superfluous power consumption in localized regions [49].

9. The integration of renewable energy sources, such as solar or wind power, into ORAN networks has the potential to effectively mitigate their carbon footprint by partially or completely offsetting power usage [50].

More specifically, in [51], a power model was proposed to evaluate the PC based on the network loads. The power model pointed to a nonlinear variation of the PC while increasing the load, this model was also suggested to beyond LTE networks. In [52] the ORAN architecture was suggested to execute the network functions in virtualised CUs and DUs using general purpose CPUs to create a processing pool. This pool may then be distributed to various geographical networks and have specific processing capacity, which impacts network energy consumption and performance. In [53] the study has offered an exhaustive review of various PC models, including virtualised and non-virtualised servers and data centres. These models have been categorised as invasive, machine-learning, and software-based. In additional. Intrusion-based models require the installation of invasive tools and events counters, which make PC measurement costly and complicated. Software-based models require an additional software to operate; this method is also powerhungry and complicated. Algorithms for machine learning are based on heuristics, although this method is time and power intensive.

In [54] virtualizing the core network has resulted in a huge reduction in the total PCs. A parameterized PC power is provided to investigate the different components of the cloud radio access network based on the number of VMs. The model assesses the PC and tradeoffs of a server undergoing virtualisation. Using differentiating characteristics, such as the amount of bare-metal BBUs and the number of VMs. In [55, 56], the Energy Aware Radio and Network Technologies (EARTH) PC assessment project presents parameterized PC models. These have been used to linearize the SotA BSs' PC (i.e. macro, micro, picoetc.). The condition under which the PC of the base station is proportionate to the provided bandwidth. Nevertheless, these models were unable of estimating the number of PC in future and hybrid networks, such as virtualised and ORAN networks. In [57], the investigation on the level of power reduction that may be attained by deploying CRAN instead of conventional BSs. In [58]. A CRAN PC model based on SDN is analysed based on the processed bandwidth. In [59] experiments are conducted to determine the impact of virtualisation on the PC of a single server while running certain packages and applications. This work considers a single-server case study, but does not provide a mathematical model-based platform for measuring the PC's components or system level, similarly in [60, 61].

One notable constraint observed in the majority of PC models discussed in the literature pertains to their specificity in terms of case selection, data type, and network configuration. Furthermore, the provided models are developed using exclusive intelligent software that is comprehensive and tailored to specific platforms, making it often inaccessible on an on-demand basis. Consequently, the utilisation of mathematical models that are both straightforward and capable of satisfying the needs of the general reader is the most effective approach for assessing the performance of power consumption. In general, there are some major falls within the existing works, first they are not suited for the ORAN design. In addition, some of the literature is based on the functions of each component, which leverage complexity. The other type are mathematical based models, but the model has no upper limits when the server is overloaded, such as in [62] and [63]

2.2 Quantum networks

The exciting world of quantum networks is a fascinating part of the field of quantum information science, which serves as a topic for both research and development. The transmission of quantum information, also known as quantum bits (qubits), over vast distances is an integral part of the protocols that underpin quantum communication. Quantum networks have the potential to completely revolutionise not only secure communication but also distributed quantum computing and the fundamental underpinnings of quantum mechanics. The following items are some of the most fundamental components of quantum networks:

1. Quantum entanglement distribution is going to be the focus of the conversation. The propagation of entangled qubits from one remote node to another can be facilitated via quantum networks. Entanglement is a peculiar quantum phenomenon that can be identified by the formation of correlations between different particles. These correlations lead to a state of interdependence that is unaffected by the physical distance that separates the particles. It is possible that this technology will make secure communication easier to achieve and permit the spread of quantum keys [64].

2. Quantum repeaters play a crucial role in expanding the reach of quantum communication by reducing the impact of signal loss and decoherence. Quantum communication systems function by partitioning the communication distance into smaller sections, use entanglement switching and quantum error correction techniques to connect these sections and maintain the integrity of the quantum information. Quantum repeaters utilise methods such as heralded entanglement generation, entanglement purification, and quantum error correction to facilitate the dependable transfer of quantum bits across extensive distances. This capability is crucial for constructing extensive quantum networks and enabling applications like highly secure quantum communication and distributed quantum computing [65].

3. In quantum communication systems, devices known as quantum repeaters are deployed to extend the range of quantum signals so that more users can receive them. The control of quantum information loss across extended distances, which is principally caused by the processes of decoherence and attenuation, is one of the most significant challenges faced in the field of quantum networks. Quantum repeaters are protocols or devices that have been designed expressly

to extend the range of entanglement and offer reliable quantum communication over enormous distances. Quantum repeaters can be used to increase the range of entanglement. This is accomplished by breaking the transmission down into its component components and creating entanglement amongst the nodes that serve as intermediaries [66].

4. Teleportation at the quantum level is a phenomenon that can be studied under the umbrella of quantum physics. As was said earlier, quantum teleportation is a set of operational guidelines that makes use of the phenomena of entanglement in order to transmit the quantum state of a particle to a location that is physically distant. Teleportation can be used in quantum networks to communicate quantum states from one node to another without the need for the qubits to be physically moved. This capacity is made possible by the fact that quantum networks have the ability to use teleportation [67].

5. Quantum cryptography is a subfield of cryptography that focuses on the design and implementation of cryptographic protocols that make use of concepts derived from quantum physics. This subfield of cryptography was first developed in the 1990s. Quantum networks have the potential to facilitate communication that is extremely secure thanks to the utilisation of quantum cryptography. The no-cloning theorem and the uncertainty principle are two of the fundamental rules of quantum physics that are utilised by quantum cryptography in order to ensure the integrity of the protocols used in the field. These ideas offer a built-in defence against listening in on private conversations [68].

6. The term "distributed quantum computing" refers to the process of carrying out computational endeavours by utilising a system that is comprised of multiple quantum computers that are linked together. Distributed quantum computing is a type of computing that could be provided by quantum networks. In this type of computing, individual quantum processors located in different nodes of the network work together to solve complex computational problems. This strategy

has the potential to make distributed quantum simulations, optimisation jobs, and other applications that need a significant amount of computational power easier to carry out.

7. The idea of a "Quantum Internet" refers to a fictitious network that would make use of the principles of quantum mechanics to facilitate safe and effective communication between quantum devices. This would be possible through the use of a "Quantum Internet." The creation of a global network that paves the way for quantum communication and processing on a massive scale is the ultimate goal of quantum networks. This internet-like network would connect all of the world's computers. For this to be possible, links would need to be made between quantum processors, quantum memories, and quantum communication devices in a way that is not only consistent but also scalable [69].

8. In spite of the extensive research that has been done on the theoretical underpinnings of quantum networks, there is still a considerable obstacle regarding the deployment of these networks in practise. This is mostly attributable to the delicate nature of quantum states as well as the necessity of using high-quality qubits. Despite this, there have been major developments in quantum communication and the spread of entanglement over extremely small distances.

Quantum networks hold a tremendous amount of potential for revolutionizing a variety of different fields, including the world of large-scale quantum computing and the field of encrypted communication. However, the creation and maintenance of these networks also involve considerable technological and engineering challenges. It is possible that the continued development of the field of quantum networks may result in substantial breakthroughs, open the door to novel applications, and stimulate the emergence of novel communication paradigms. These outcomes have the potential to be achieved simultaneously.

In [70], the adaptive cost that is originated from using quantum technology in classical communication has been discussed. The cluster head selection policy is solved by using the quantum approximate optimisation algorithm to achieve an energy-efficient network.

In [71], the technical aspects of quantum computer based systems such as quantum memory, quantum gate, quantum control, and quantum error correction have been introduced. The entropy of quantum channels is studied in [72]. In [73], a quantum repeater was proposed to reduce network errors while evaluating the channel capacity. In [74], a satellite has been utilised to exchange entangled photons over one hundreds of kilometers channel long. In [75], as well as in [76], models are proposed to provide a solution by using entanglement security in quantum internet networks. Moreover, the authors of [77] have proposed a multi-layer process for optimising internet based quantum networks. This technology limits the processing time of the node's quantum memory, improves the connection performance, and reduces the amount of signaling. In [78], entanglement theory is used to protect network security by enabling quantum based key distribution. Following, the researchers in [79] have used the free space to distribute entangled photons over 13.5 km experimentally. The authors showed that these photons can always survive such a long distance. Subsequently, classical data is transmitted between parts through quantum teleportation channels [80]. In [81], the progressive bits are encoded using optical fiber by using a transmission connection, and the transmitted photonic array is used to improve the final network throughput. The author of [82] showed that traditional data and quantum data could be transmitted cooperatively.

In [83], the authors have distributed high-dimensional quantum states over 2 km of multi core fiber. They demonstrated how their implementation would benefit from quantum bits' advantages, e.g., their higher noise resilience and greater information power. However, in [84], it was found that the communication costs in quantum networks are at least twice the cost of traditional networks using the same number of parameters. Furthermore, Table 2.1 provides updated protocols and improvements related to quantum communications.

Quantum mechanic can boost the performance and security in several technologies, such as quantum sensing [85], quantum metrology [86], navigation and timing [87], state discrimination [88], quantum communications [89], and computation [90]. These technologies started as theoretical marvels but have made significant progress, and are now widely applicable. Improved quantum sensors are enabling the first mainstream use of quantum technologies. Quantum correlations in probe states provide precise limits for physical signal measurements [91]. QKD methods have also accelerated secure quantum communications [89]. Early adoption of specific quantum technologies has increased interest. Concentrated efforts have improved performance across all quantum technologies, but they have not met the technological need. Most applications require numerous quantum technologies. Quantum entanglement, the main resource in networked quantum technologies, can be disseminated directly [92] or via photons swapping [93] and purification [94] actions on topologies of quantum nodes that can be completely arbitrary [95]. Two quantum entanglement features guide networked quantum devices. Quantum connections between entangled nodes initially allow remote processors to coordinate. This simplifies quantum clock synchronization [96]. Quantum communication was developed because quantum computers threatened crypto-systems. Quantum computers are more powerful than classical computers and can use publicly available quantum processors remotely. High-speed global quantum communications networks are needed for delegated quantum computing. Networked quantum computation and quantum communications are shaping a quantum Internet [97]. Like the Internet, this is anticipated to change our world. Quantum networks' distance-limited range precludes this idea's implementation. Quantum communications research is increasingly aimed towards circumventing this constraint. Photonic losses doom the optical fiber only quantum Internet. Entanglement routing using quantum repeaters improves this capability [98]. Quantum networks have established connecting major cities, such as in China [99].

Some other works can be found in Table 2.1, which shows some methods and applications related to quantum communications and quantum computing.

Table 2.1
Related works for quantum computing.

Method	Applications	Research
quantum networking	wireless communications	[100]
private quantum	mobile communications	[101]
super-dense coding	decoding the quantum bit	[102]
non-cloning	ciphering	[103]
compression	coding	[104]
entanglement	quantum broadcasting	[105]
optical communications	communication protocols	[106]
key distribution	quantum security	[106]
unique numbers generation	quantum coding	[107]
channel capacity	quantum channels	[108]
concentrating	entanglement transformations	[109]

2.3 load balancing and optimisation

When it comes to ORAN, the usage of load balancing schemes is an extremely important factor in the optimisation of resource utilisation, the improvement of network performance, and the preservation of a consistent QoS for UEs. Because ORAN architectures typically includes features that are centralized and virtualized, the deployment of load balancing algorithms is required in order to properly distribute network traffic over a variety of network components. Within the scope of ORAN, it is possible to single out a number of load balancing strategies that are currently in use.

1. User association and cell selection:

-Proximity-based association: The method of associating UEs with the base station or cell that is physically located the closest to them is utilized in order to reduce the amount of signal loss that is caused by path distance.

-The load-based association: The method involves assigning users to cells based on the load that is now being experienced, with the goal of fostering more distribution of the resources [110].

2. Frequency and spectrum management:

   -Frequency load balancing: The practice of effectively distributing frequencies and spectrum resources to cells or sectors is carried out with the intention of lowering the amount of interference and congestion that occurs.

   -Dynamic spectrum sharing: It refers to a method that allows for the dynamic allocation of spectrum to individual cells. This method takes into consideration the fluctuating demand for spectrum. This strategy makes the most of the spectrum resources that are at our disposal, which results in an increased level of spectrum utilization efficiency [111].

3. Traffic steering and steering policies: This means directing specific type of traffic or users into certain cells or paths, while taking into consideration the QoS requirements and the conditions of the network. Subsequently, policy-based steering is the practice of utilizing established policies to determine the routing of network traffic. This approach is also known as "steering based on policies." These policies are decided upon after taking into account a number of factors, such as the type of application that is currently being utilized, the intended levels of latency, and the priorities of particular users [112].

4. Dynamic resource allocation: is the process of allocating resources in a network such as bandwidth, power and processing capacity depending on real-time network situations and traffic demands. This process is referred to as the method of dynamic resource allocation [113].

5. Cloud and edge load balancing: The allocation of edge computing resources is balanced to effectively share processing and storage needs among edge servers. In centralized cloud computing, the degree of demand serves as the determining factor in the distribution of central cloud resources [114].

6. Traffic Offloading and Redirection: The process of transferring network traffic from one network node to another in order to alleviate congestion and maximize network performance is referred to as the idea of traffic offloading and redirection. The term "traffic offloading" refers to the process of rerouting traffic to different paths, cells, or network elements in the event that congestion or network problems are detected. Subsequently, the term "redirection policies" refers to a collection of guiding principles that are established in order to make the process of rerouting network traffic easier. These policies are intended to establish the right course of action for redirecting traffic based on a number of criteria, some of which include, but are not limited to, load distribution, latency, and specific service requirements [115].

7. Machine learning and artificial intelligence: these are two of the components that load balancing algorithms make use of. In order to assist judgments regarding load balancing, techniques from the field of machine learning are used to examine both historical data and the conditions of the network in real time. Subsequently, artificial intelligence employs complex decision-making systems to dynamically optimize load balancing in order to achieve optimal performance [116].

8. Network slicing: is the act of logically splitting networks into separate slices, each of which is designed to cater to a certain set of services or applications. Network slicing is also referred to as "virtualization" This makes it possible to apply individualized load-balancing algorithms within each unique slice of the system [117].

The coherent approach is guaranteed by the incorporation of a variety of techniques in network management, which ensures alignment between various areas. Real-time adjustments are facilitated by the monitoring and adaptation of network conditions and performance metrics in real time. AI-driven optimisation continuously learns and enhances decision-making processes by analysing data across all techniques to find the optimal balance. Traffic steering, resource allocation, and redirection policies are enforced by policy enforcement to ensure that they are consistent and that they meet the quality of service (QoS) requirements. Network management can achieve high efficiency, flexibility, and user satisfaction by integrating these techniques, which enable it to adapt to altering conditions and demands in real-time.

A number of factors, such as the network architecture, the deployment environment, and the particular QoS needs, all play a role in determining whether or not adequate load balancing solutions can be successfully implemented. By utilizing a combination of these tactics, it is possible to accomplish efficient load distribution as well as network optimisation, which will ultimately result in an improved experience for the end user.

Given the inclusion of load balancing and optimisation problem in ORAN as a primary focal points, it is uncommon to encounter comparable studies in the context of ORAN. Consequently, the existing literature is segregated into several categories corresponding to these themes. The authors' objective was to investigate the most closely related research to the subject under consideration.

Regarding ORAN, there are still ongoing technical inquiries that seek to tackle its simultaneous obstacles and resource allocations. The existing body of study is constrained in its scope to encompass the examination of several aspects such as the difficulties encountered, progress made, structural framework, valuable perspectives, and proposed remedies. For instance, in [118], the ORAN architecture is presented to facilitate multi-vendor interoperability by utilising disaggregated, virtualized, and software-based components that are interconnected through open and standardised interfaces. The conventional RAN services are commonly characterised by their private and closed nature, leading to elevated expenses and a dearth of transparency for network operators. To effectively manage the limitations and minimise expenses, it is crucial to initiate a network redesign process that focuses on improving the efficiency of RAN installations and streamlining the operational aspects of RAN network services, as proposed in the study by [119]. The authors of [120] proposed an end-to-end network slicing method in multi-cell system. Nevertheless, the application of this technology in the RAN continues to provide significant challenges. In the study conducted by [121], it was shown that ORAN adopts cloudification and network function virtualization as methods for processing baseband functions. it showed that using heuristics enhances the economic efficiency and optimises the network resource in comparison to traditional greedy resource allocation algorithms. Furthermore, in [38], the authors shed light on the existing constraints of the present ORAN standards and proposes potential technological solutions to address these restrictions, while the presented study of [122] focused on the development of a near real-time RAN intelligent controller service by the open networking foundation. The study also presents simulation findings pertaining to this service. Trends and opportunities are discussed in [39], advances in [123], and programmability of ORAN in [37].

In accordance to load balancing, the study conducted by [124] focused on the development of load balancing techniques within the RAN in the context of network slicing. In the study conducted by [125], an optimisation problem was presented with the objective of selecting the optimal split points for the ORAN. The primary aim is to achieve load balancing across CUs and midhaul links, while also taking into account the delay criteria. The formulation that arises from this problem is classified as NP-hard, and it is addressed using a heuristic algorithm. In a previous study [126], a load balancing technique was introduced with the aim of improving the overall sum-rate performance of the ORAN. Two sub-approaches were described that have the ability to function independently in a non-realtime RAN intelligent controller and a near-RT RIC, respectively. The findings demonstrated an improvement in the effective network sum-rate, along with enhanced load balancing among the radio units. A reinforcement algorithm was proposed in [127] for ORAN radio intelligent controller. load balancing in cloud radio access network can be found in [128], [129] and [130]. However, neither of these studies incorporated a quantum domain in their work, nor did they address the EE of ORAN.

Chapter 3: Power Consumption Evaluation of Next Generation Open Radio Access Network

ABSTRACT

ORAN executes network functions in three types of units: central units (CU), distributed units (DU), and radio units (RU). ORAN has evolved as a tool to deliver spectrum sharing and cost-effective solutions. However, this will result in a rise in network infrastructure and power consumption (PC). To assess the PC, a model is required to measure the trade-offs of a server undergoing virtualisation in an ORAN infrastructure. Different parameters, such as the number virtual machines (VMs), system's resource blocks (RBs) and bandwidth have been used and compared with bare virtualization and traditional cloud networks. The term bare metal refers to the fact that there is no operating system between the virtualization software and the hardware. Virtualised network (VRAN) has resulted about 50% reduction in the total PC in comparison with traditional cloud networks. However, the ORAN has produced more PC compared to VRAN, about 30% in the total PC and 10% in the cooling PC. Unless the advantages of ORAN are fully utilised, the addition of extra units within the ORAN, specifically the DU servers, may result in more PC that might advocate against the ORAN.

3.1 Introduction

Because of the increased number of users and network devices, mobile operators and equipment suppliers have prioritised adopting the open RAN (ORAN) concept due to the necessity to deliver at least 10 times more spectral and energy efficiency in 6G networks [63,131,132]. In ORAN, networks are constructed with completely open interfaces, protocols, hardware and software, that operate on commercial, off-the-shelf (COTS) servers [133]. Mobile networks have always been built using closed, proprietary software and purpose-built hardware. However, it may now be decentralised and based on the ORAN concept. In this instance, this refers to the separation of hardware and software. This concept was introduced in Release 14 of the 3GPP standards, which separates the control and user planes of evolving the nodes, whilst developing the ORAN specifications [134]. In 3GPP Release 15 and beyond, the service-based architecture continues this separation with the development of distributed techniques such as the virtualised RAN (VRAN) and edge computing. Consequently, the RAN functions are divided into a number of modules that can be distributed across different units, possibly located in distinct locations: the RU located as close to the antenna as possible, the DU located further away, and the CU at the edge of the network, as presented in Fig. 1. [135]. It is essential to recognise that virtualisation and ORAN are not synonymous. Virtualisation is the separation of hardware and software by decoupling the application software from the hardware that runs on, while ORAN means that the hardware is virtualised and these virtualised softwares may serve different network vendors and operators [38].

Recently, the research community has embraced the use of network function virtualisation (NFV) techniques in the cloud for a variety of reasons including [62]: flexible allocations for the available network resources, enabling automation in the servers' operation and configuration, reducing the cost of maintenance, and saving more power. NFV enables the operation of fewer processing units in the cloud while still meeting the quality of service requirements of the users (UEs) by installing many virtual machines (VMs) on a single server, each operate as an individual server [136] [137]. However, in order to operate several VMs on a single host server hardware, a supervisor or manager is required. This is also called as a hyper-visor (HV). It is a software that dynamically control and share the host with guest VMs. Then, each VM seems to have exclusive use of the server's memory (RAM), processors (CPU), network interface card (NIC), and hard drive (HDD). However, each VM shares these resources with other VMs. The HV ensures that the hosted VMs may not interfere with one another while accessing these resources.

However, a RAN may be virtualised but not open, meaning that the software and/or hardware may be proprietary and/or the interfaces may be closed. Being completely "open" necessitates the existence of reference designs and standards for hardware and software, so that the RAN has only open interfaces and no proprietary interfaces and/or hardware [138, 139]. In which, a radio head (RRH), or also called radio unit (RU) belongs to operator/vendor A is able to communicate with (proprietary) software operating on a COTS server from vendor/operator B through open interfaces. Note that "openness" does not imply that all hardware and software for mobile networks will be similar. Vendors will compete to supply all hardware and software so that operators have a diverse range of options in terms of size, scope, features, and price, yet these devices are compatible and adaptive with any open software. Virtualization has the potential to offer legacy vendors/operators a number of benefits. They are not required to adapt their hardware or software interfaces with other vendors, but they are allowed to comply to the 3GPP release requirements [140, 141], yet, their hardware is proprietary built. Therefore, they continue to push and supply only closed, proprietary technologies that serve only their best interests and do not establish a future-proof network for their clients, despite the fact that doing so relies within their capabilities [142]. One of the motivations for ORAN has been the need to extend the radio system, since there are currently relatively few radio vendors. One of a mobile network's most important elements is the radio interface since it connects the UE to the network. It was believed that expanding the radio environment is essential for the success of the ORAN evolution [143–145]. In contrast to current cellular networks, ORANs are remarkably capable of adopting cooperative algorithms, dynamically using the available spectrum, leveraging load variation to run less computing resources, and integrating with the latest 6G enabling technologies with least cost [35, 146]. It is worth mentioning that the ORAN network include separating the legacy BBU hardware by proposing central unit (CU), connected to distributed unit (DU), the latter is connected to the RU. The DU is responsible for MAC/RLC and High-PHY processing, while the CU is responsible for RRC/Control plane processing and Transport/S1 [140]. Nevertheless, increasing the number of deployed open devices, changing the network architecture design and installing new protocols may sustain a substantial amount of PC, that has to be estimated [147]. There are many reasons that causes the PC variations: installing more softwares within the server, adding new devices to the network, removing/adding some functions from/to the server, increasing/decreasing the number of UEs allocated to each software, increasing/decreasing the number of resource blocks of each UE, etc [57]. This chapter then establishes a model to calculate the PC of the ORAN while considering the above network parameters, such as bandwidth, number of VMs, modulation and coding schemes (MCS) and number of UEs. The proposed model is mathematically-oriented, easy to follow, and represents a general tool for the researchers that require a power model in their evaluation as a main or side measurement. It further defeats the complexity of other models that are based on the functions of each units, rather than the network parameters. As far as the authors know, there is no power model that estimate the PC of the ORAN network.

3.2 Contributions

1. There are huge demands in the next generation's specifications pushing towards reducing the PC and carbon footprint by reducing the PC. This cannot be achieved unless PC model is existing to measure the reduction for the long range scenario, and the evolved network architectures.
2. The proposed model is different than the other models that are complex and functions based. It is mathematical-wise that is simple to use. Yet, it maintains the accuracy level by considering the most effective parameters such as bandwidth, resource blocks,number of VMs and MCS type.
3. The researchers that propose PC reduction strategies, algorithms, optimisation techniques and machine learning based methods require simplified and easy to adapt PC model to evaluate their results. The proposed model facilitates such enquiry.
4. A comparison amongst the PC of ORAN with cloud radio access network and VRAN has been accomplished to show the effectiveness of the model.

3.3 PC model

When developing 5G, the 3GPP first took into consideration the split concept of the BBU server into DU and CU. Practically, some functions of the BBU have been migrated to the DU in different servers and different locations. The other functions are left in the legacy CU, as shown in Fig. 3.1. In spite of the fact that CUs continue to have features comparable to lagacy BBUs, DUs will have great processing capacities.

3.3.1 CU model

The CU is able to perform a wide variety of functions, including broadcasting information, establishing and releasing connections to user equipment, data transfer,

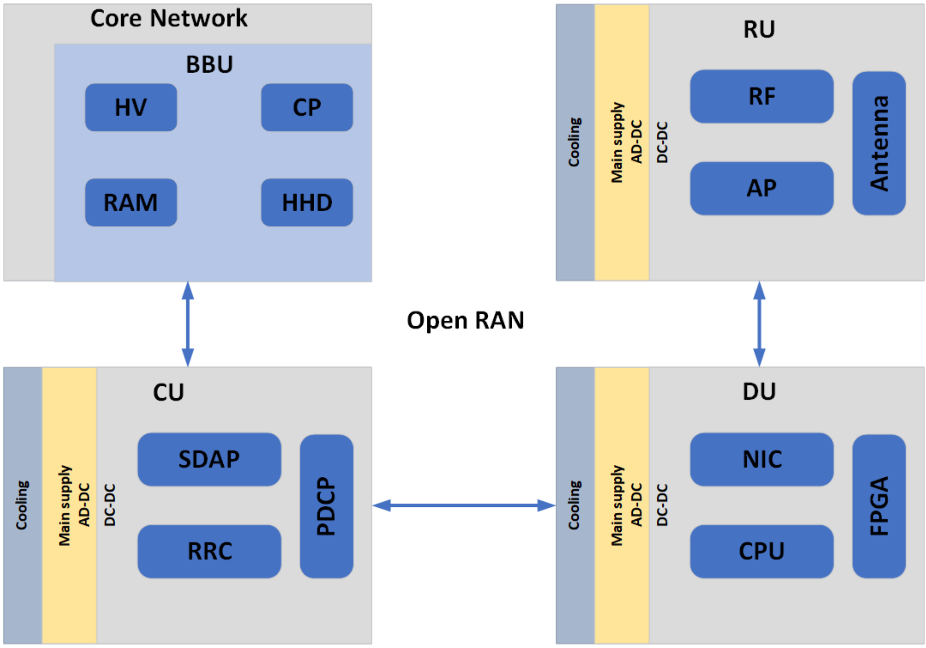

Fig. 3.1. Block diagram of ORAN power component

performing quality of service functions, and compressing and decompressing IP data streams. The PC of the CU is typically calculated as the total sum of active CUs included within the CU pool. The CU's digital computation and processing may be evaluated by Giga operations/second (GOPS), and this metric can then be translated into power [148].

It is possible to link GOPS with a collection of multiple CU's functions. However, this representation is complex and intractable due to its high subsequent scaling, parameters and function's measurement. However, it is possible to express the PC of the CU as:

$$\frac{dP_{CU}}{dBW} = \alpha P_{CU} \tag{3.1}$$

Where the change in PC of the CU server is proportional to the change of bandwidth (BW), meaning, proportional to the number of processed resource blocks, packets or coming load. This concept has been previously mentioned in [149] as the dynamic PC is based up on the load variation. After solving this equation, the result is

$$P_{CU}(BW) = P_{CU}^{initial} e^{\alpha BW} \tag{3.2}$$

Where α is increasing constant, and the $P_{initial}$ CU is initial PC of CU in idle mode of operation. In addition, the CU consumption is affected by the type of modulation and coding scheme. Using the same style, we can formulate the change in the PC is proportional to the change in the modulation and coding scheme, denoted as modulation coding scheme (MCS), which is usually describes the number of bits transmitted within each resource element. This means the higher MCS, the more data are transmitted.

$$\frac{dP_{CU}}{dMCS} = \lambda P_{CU} \tag{3.3}$$

When this equation is solved, it yields:

$$P_{CU}(MCS) = \lambda P_{CU}^{initial} e^{\lambda MCS} \tag{3.4}$$

The total CU server consumption P_{CU}^{T} is the joint addition of the BW and MCS effects, as follow:

$$P_{CU}^{T} = {P_{CU}}^{initial}(e^{\alpha BW} + e^{\lambda MCS}) \tag{3.5}$$

It is worth mentioning that the ORAN only has bandwidth and MCS effects, but the virtualisation of the CU also increases the consumption when several VMs are found and compete within the same server, for example, in the CPU, they boost its PC as each VM has its own packets to process. Hence, the effect of virtualisation can also be added to the total consuumption of the $CU P_{CU}^{Pl}$ as follows:

$$P_{CU}^{Tl}(BW, MCS, N) = {P_{CU}}^{initial}(e^{\alpha BW} + e^{\lambda MCS} + (e^{\alpha N}) \tag{3.6}$$

Where N is the number of VMs. The Hypervisor also consumes an amount of energy within the server. If we assume the PC of one task (t) per VM_n is $P_{n,t}$, then the PC of the HV may be modeled as:

$$P_{HV} = \sum_{n=1}^{N} \sum_{t=1}^{T} P_{t,n} \tag{3.7}$$

where $P_{t,n}$ is the PC of the t-th work given to n-th VM, in addition, T is the total number of tasks. Finally, the total PC of the CU unit is updated, as follows:

$$P_{CU}^{Total} = {P_{CU}}^{initial}(e^{\alpha BW} + e^{\lambda MCS} + (e^{\alpha N}) + P_{HV} \tag{3.8}$$

3.3.2 DC-DC power consumption

In order for an electronic chip to function properly and satisfy the requirements outlined by the manufacturer's specification, the device has to be supplied with a certain DC voltage. This unit is in charge of converting high levels of DC voltage to an operational voltage. Because the DC-DC converters have a level of efficiency, its power model may be described in terms of losses. These losses are directly linear with the number of power components that use DC voltage [56]. The PC of the CU's DC, denoted by (P_{dc}^{cu}), is modeled by taking into account the DC-DC losses (ℓ_{dc}^{cu}) as a function of the efficiency (η_{dc}^{cu}). This is done in conjunction with the power requirement of every other component.

$$P_{dc}^{cu} = \ell_{dc}^{cu}(\eta_{dc}^{cu})[P_{CU}^{Total}] \tag{3.9}$$

3.3.3 AC-DC power model

This unit is in charge of converting the power coming from the mains supply grid from alternating current to direct current. When modelling the overall PC of AC-DC (P_{ac}^{cu}), the same way is used as when modelling the DC-DC power conversion. This is modeled thus in order to take into account the losses that occur during the AC-DC conversion. It is possible to model the PC of the CU's AC conversion in CU, by taking into account its losses, which are denoted by the symbol (ℓ_{ac}^{cu}), as a function of the efficiency, which is denoted by the symbol (η_{ac}^{cu}). This is done while simultaneously taking into account the power requirements of each and every component. The power model is able to be summarised as follows:

$$P_{ac}^{cu} = \ell_{ac}^{cu}(\eta_{ac}^{cu})[P_{CU}^{Total}P_{cu}^{dc}] \tag{3.10}$$

3.3.4 Cooling power model

A PC model for a CU that needs cooling may be modeled as a constant power loss that scales linearly and in proportion to the amount of power consumed by the unit. If ℓ_{cool}^{cu} represents the consumption of cooling loss in the CU, then P_{cool}^{cu} may be derived as the amount of cooling power consumed.

$$P_{cool}^{cu} = P_{ac}^{cu}\ell_{ac}^{cu}(P_{CU}^{Total})(P_{cu}^{dc})(P_{cu}^{ac}) \tag{3.11}$$

Finally, the total PC of the CU unit P_{CU}^{T} is as follows:

$$P_{CU}^{T} = P_{CU}^{Total} + P_{HV} + P_{ac}^{cu} + P_{dc}^{cu} + P_{cool}^{cu} \tag{3.12}$$

3.4 Distributed unit

The DU is a COTS edge layer server that may function as a BBU to manage high layer PHY, MAC, or RLC functions [140]. First, a VM layer software may be installed to serve several RUs. Subsequently, these VMs make it possible to fully utilise the server's resources. On top of such layer, the DUs software that the vendors provide may be installed and controlled by the HV, similarly to the CU unit. By reducing raw performance, reducing the required number of CPU cores, and lowering PC, hardware acceleration has the potential to significantly enhance the efficiency of any ORAN [150]. When the value of N is increased, the PC of each DU server increases. Because of this, the model has to include a description of an increase in the PC in order to allow the virtualised server to be driven to achieve its maximum PC. However, the PC of this unit is modeled the same way as the CU unit, as follows:

$$P_{DU}^{T} = P_{DU}^{Total} + P_{HV} + P_{ac}^{du} + P_{dc}^{du} + P_{cool}^{du} \tag{3.13}$$

3.5 Radio unit

The RU converts radio signals from/to the antenna into a digital signal that may be transferred through the fronthaul from/to the DU, the RU consists of the following components:

3.5.1 RF unit

The RF transceiver unit mainly comprises of an interface with intermediate frequency and base band modulation/demodulation. RF unit is accountable for several functions that are described in [57]. Its PC was modeled as a constant value that 12.9 W, which was the simplest way to model such consumption. However, in [151], it was mentioned that RF PC is slightly influenced by the bandwidth. Hence, RF's PC (PRF) is modeled as a load affected unit that its PC increases when more network UEs are connected. Here the linear model was used to describe RF PC, as follows:

$$P_{RF} = BW \times P_{RF}^{int} \tag{3.14}$$

It is worth mentioning that the value of the bandwidth cannot be multiplied directly to the initial PC of the RF, otherwise, the PC will be boosted greatly. Hence, it is preferred to normalise the value of the bandwidth or convert its value to a normalised number of resource blocks to obtain valid outcomes.

3.5.2 Power amplifier (PA)

The PA is a primary concern in the RU since it uses the majority of the RU's power. The electrical signal that was recovered from an O/E converter is amplified by the PA before being delivered to the air interface through the antenna, and vice versa. The PA usually has low efficiencies at low transmission power (P_{out}); nevertheless, if significant transmission power is desired at the antenna, its efficiency may reach up to 54% or less due to the significant variation in transmission power of OFDM signals [152]. The amplifier PC, denoted by P_{amp}, is influenced by the its efficiency, denoted by ηamp, which is a function of the transmitted power of the antenna P_T. The P_{amp} might be modeled as

$$P_{amp} = \frac{P_T}{\eta_{amp}(P_T)} \tag{3.15}$$

The PC of RU is the addition of the amplifier and RF unit:

$$P_{RU} = P_{rf} + P_{amp} + P_{dc}^{ru} + P_{ac}^{ru} \tag{3.16}$$

Where P_{dc}^{ru} and P_{ac}^{ru} denote the PC of the DC-DC and ac AC-DC, respectively. These powers are modeled the same style of the CU or DU units. It is worth mentioning that the RU unit does not include cooling unit because its low PC. Therefore, the total PC of the ORAN (P_{ORAN}) is expressed as follows.

$$P_{ORAN} = P_{CU}^{T} + P_{DU}^{T} + P_{RU} \tag{3.17}$$

3.6 Results

To correlate the findings of this work with a real-time measurements, the resulting parameters were selected from [54, 57], as shown in Table 3.1. We have assumed the number of CUs, RBs and VMs are up to 50. However, because of how easily the model may be adapted to new circumstances, this number is subject to change. It is possible to replicate the configurations of the servers sold by a variety of vendors by changing the values of the different parameters and observing how the end model varies as a result. For example, the initial PC of some servers are different than others, this means these results vary, the new results can be simply produced once applied to the model. Moreover, the maximum PC of the server is not the same for all of them, this means changing the tuning parameters, such as $\aleph$ and λ, is necessary to control at what number of VMs the server reaches the maximum power. Another obstacle is that this model assumes that 50 VMs drive the server to its maximum power, where in practice, may be not. The servers' behaviour to the VMs is different, hence, it is required to physically measure the maximum PC and compare it to the data sheet of the server to estimate the exact number of VMs. After, the model parameters can be easily adjusted to such a specific type of server.

Table 3.1

Model Parameters I

Component	**Unit**	**Value**	**Component**	**Unit**	**Value**
P_{RF}^{int}	W	12.9	$MCS, 16$	W	9.5
$P_{CU}^{initial}$	W	26.5	α	-	0.003
$P_{DU}^{initial}$	W	29.6	η	-	0.008
P_{amp}	W	6.1	λ	-	0.1

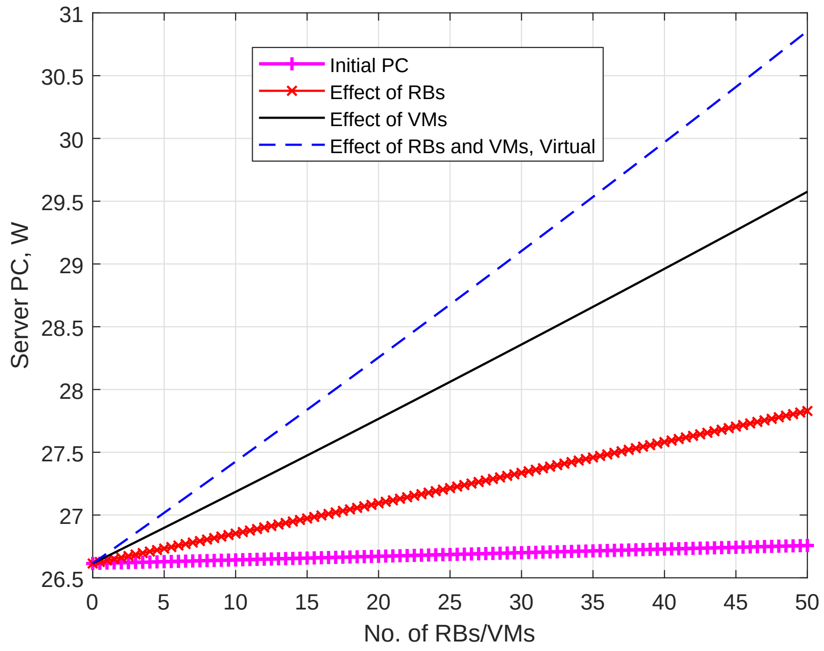

Fig. 3.2. The PC of a server showing the effect of the number of RBs or VMs.

Fig. 3.2 shows the effect of RBs and VMs up on the server PC, the latter drives the PC of the server to its maximum. However, the number of VMs can be reduced to as many as practically installed VMs while keeping the maximum PC the same. The model allows to tune such case by using the model's different parameters. It was assumed that each VM runs up to 50 RBs.

Fig. 3.3 shows the PC comparisons of the ORAN with CRAN and VRAN while running 10 and 20 virtualised CU servers in the CU pool. The highest two values show the initial PC of the pool without virtualisation, i.e. CRAN, with and without the effect of RBs. The number of RBs is assumed equal to 50. Meaning, that each VM is responsible for processing 50 RBs. However, as the number of RBs increases, the PC will be increased too in the CU pool. Subsequently, the other indicators show the PC comparison of the virtualised and ORAN case. Yet, the ORAN has produced more PC compared to VRAN, about 30% of the total PC. This is due to

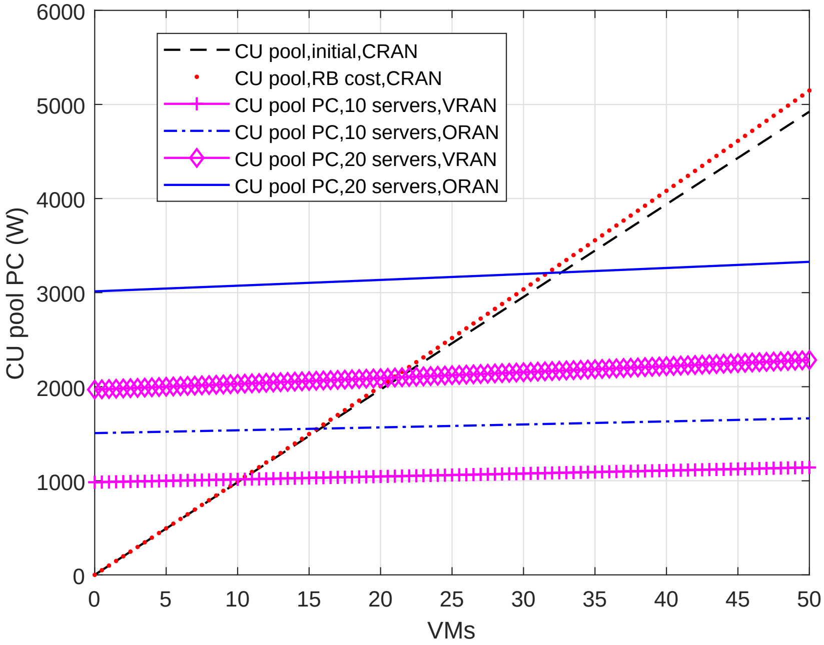

Fig. 3.3. Comparison of total PC of ORAN with CRAN and VRAN.

the addition of the DU server to the ORAN architecture. Moreover, the DU itself is also virtualised. In contrary, the VRAN case uses only one server, that is BBU server. Both ORAN and VRAN show less PC than the traditional architecture, where no virtualisation can be found. Although the PC of ORAN is more than the VRAN. However, the sharing nature of the former can manipulate such matter as this cost can be divided/shared amongst the sharing network operators that utilise the server resources. Not to mention the delay cost. However, in the pure VRAN, only one network operator endures the complete cost.

In terms of accuracy, the results of the model will be affected in a manner that is proportional to the degree to which each piece of equipment maintains unique operating conditions. These conditions include initial and maximum PC, cooling requirement, and efficiency. However, the model can be relied upon to accurately appraise the PC since it is based on concepts and assumptions that are derived from

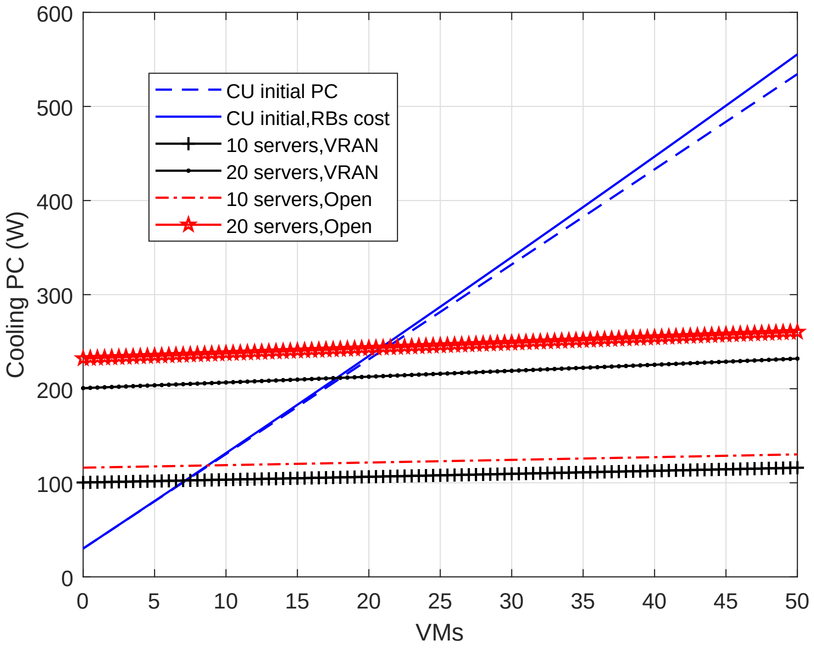

Fig. 3.4. Comparison of cooling PC of the CN with, without virtualisation and ORAN of 100 BBUs or VMs while processing 100 RBs.

actual data measurements. Subsequently, Fig. 3.4 shows the cooling PC comparison of the CRAN, VRAN and ORAN when 50 VMs are hosted by 10 and 20 servers. In both cases, this is due to the cooling requirements at the DU and CU servers. The ORAN shows about 10% more PC than VRAN, however, less than CRAN.

3.7 Conclusions and future aspects

In this work, the cost of increasing the number of VMs and the number of RBs on the ORAN servers is evaluated. In addition, a comparison for cooling and overall PC of ORAN with CRAN and VRAN is presented. The proposed model is flexible enough to accommodate the changing in the values associated with any type of servers. Intuitively, virtualisation results in a decrease of network PC compared to the traditional CRAN. However, due to the addition of the DU servers within the

ORAN, it scores more PC compared to VRAN, about 30% in the total PC and 10% in the cooling PC, but far less than the CRAN. However, this study has accomplished the PC comparisons without considering the multiple vendors gain in terms of PC and delay. Moreover,the PC increment in ORAN case can be compromised by the gain of DU servers, where DU functions become closer to the user, which lessen the end to end delay. Another gain can be added to the ORAN when measuring the gain of each vendor individually. Finally, there are new advantages that can advocate the ORAN, such as sharing the resources, sharing the power cost, inter and intra servers load balancing.

Chapter 4: Quality of Service of Quantum Entanglement in Mobile Networks

ABSTRACT

There are many problems in cellular communications that cannot be resolved traditionally. The quantum communications can add new dimensions, safety, encryption and solutions to the traditional networks because of its robust physical strength. However, it is not entirely realised how to adapt the quantum into traditional communications because it is not entirely utilised. This chapter addresses the necessary guidelines and assessments for future quantum solutions to the standard mobile cloud networks. In particular, using entanglement phenomenon to increase the performance of the X2 application (X2-AP) protocol by minimising the overhead signalling, represented by the time and energy consumption the conventional cloud encounters. We intended to offer a delay reduction while adapting the quantum technique into the cloud by modelling the latency of both paradigms. Finally, increasing the number of photons has decreased the delay to about 40% compared to the traditional network. In addition, the energy efficiency in the quantum case has been increased while decreasing the power consumption by about 10%. The application of quantum entanglement concept can lead to reduced power consumption by reducing signalling overhead. This is based on the assumption that this technology is currently available, with a mathematical framework providing a model for its behavior.

4.1 Introduction

Some work has been done to maximise the usable bandwidth resource blocks. However, occupying the stingy amount of resource blocks shall come to an end due to the inherent low available bandwidth, and maximising these will no longer increase the spectral performance [153]. As per the definition of resource blocks, the stingy amount of bandwidth means that these resource blocks are only available in a few amounts each ms. Therefore, the struggles continue in the classical communication even when the most effective technologies are used [154]. As per the definition of resource blocks, the stingy amount of bandwidth means that these resource blocks are only available in a few amounts each ms. Not to mention the inherently unsolvable delay problem that is physically related to the distance between the transmitter and receiver, and processing delay. This causes the communication calls to be blocked and UEs outage [155]. Hence, the quantum domain may offer the required solution [156]. Generally, applying quantum methods to mobile communications is unusual. The truth is that quantum computation is incomplete itself [157]. Moreover, classical behaviors and quantum behavior vary tremendously [158]. Optical communications technologies have several quantum features represented by optical fibers, laser sources for photons to be produced, and the light on the receiver side to be sensed [159]. However, only one wave property is utilised and seen in the classical sense out of the two photon characteristics. The photon operates based on how a photon is measured and modified in both wave and particle properties [160]. Recently, quantum computing applications and advances have been spreading, such as quantum entanglement, quantum routing, quantum repeating, quantum relay and encoder/decoder, quantum synchronisation, quantum memory and quantum cryptography [161].

In the literature, the upcoming cloud networks has been a candidate for next generation, especially 6G, to reduce the power consumption of the traditional networks. By combining the base band units (BBUs) of the legacy sites in one place, leaving the cell site as simple as it contains the antenna, amplifier and radio frequency unit, called remote radio head (RRH). This requires less cooling, less total power consumption, less renting cost, and more cooperative and collaborative procedures will be gained. However, some disadvantages have been assured such as the need for more complicated algorithm to run the network. In addition, more channel delay that is originated due to shifting the data plane processing to far-away data centers. Furthermore, a more significant number of signalling control planes contributes to higher power consumption, complexity, latency and increases the rate of blocking calls [162]. In contrast, the ideal handover must overcome the traditional procedure and offers less power consumption and less delay. In this work, we proposed a power and time delay saver approach that uses quantum entanglement to reduce the inherent signalling delay of the X2-AP protocol, classically used for the handover process. For that purpose, we have proposed a time delay model to measure the classical and quantum delays. The latter has caused some power consumption and energy efficiency trade-offs within the quantum method compared to the traditional network. Nevertheless, a simplified power model has been proposed to calculate both classical and quantum network consumption. We may summarise the contributions of the proposed work as follows:

1. The already used X2-AP handover protocol causes a large amount of signalling represented by the time and power consumption [163]. Subsequently, quantum entanglement phenomena has been used as a handover process instead of the traditional method. The former can utilise a hidden channel amongst the generated photons to transfer the information amongst the mobile radio heads with zero delay. A hidden channel in quantum communication is a covert pathway used to securely transmit information, leveraging quantum mechanics to conceal the data and detect eavesdropping.

2. The proposed method has replaced the successive classical signalling, each with a corresponding entangled photon. Classically, when the remote heads try to communicate with each other asking for handover, the destination and source remote head uses classical signals with time and energy perspectives. The quan-

tum method has replaced such communication by changing the behaviour of one of the entangled photons (suppose in the source remote head) to pass the information without delay to the other photon (supposed in the destination remote head) to reduce the overall latency of the network.

3. Passing information without delay means the power consumption can also be reduced. Classically, such power consumption originates from generating the classical signalling and transmit to the other network parties. In the quantum method, the power consumption of generating the signals has been ignored. Rather, it was replaced by a consumption that is originated from generating the entangled photons, circuit drivers and receivers. These consumers have been compared with the classical method by deriving power models for both methods. Based on the latter, and the UEs data rates, the energy efficiency has been compared by assuming the network is serving an amount of UEs aimed to move from one cell to another.

As far as the authors know, there is no similar work that tackled reducing the delay and power consumption by using the quantum method that is adapted within the mobile network.

4.2 Quantum fundamentals

4.2.1 Qubit

The quantum bit, also known as a qubit, is the most fundamental component of quantum information. It is a representation of quantum state in quantum computing. The mathematical representation of the state of a qubit can be given as:

$$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle \tag{4.1}$$

Where the normalization criterion $|\alpha|^2 + |\beta|^2 = 1$ is satisfied, and α and β are complex numbers that denote the probability amplitude of the photon being in the state $|0\rangle$ or $|1\rangle$, respectively [164]. The core difference with classical bit is that the latter can be in the state 0 or 1 independently, while the qubit rises the probability of being 0 and 1 at the same time. Some of the applications of qubits include, but not limited to:

1. Quantum computing: In order to perform computations, quantum computers make use of qubits, which allow them to solve certain kinds of problems in a significantly shorter amount of time compared to traditional computers [165].

2. Quantum cryptography: Qubits can be used to transmit information securely using QKD, a method that ensures the confidentiality of the communication by utilizing the properties of quantum mechanics [166].

3. Quantum simulation: Researchers can use qubits to mimic sophisticated quantum systems, such as molecules and materials, in order to improve their understanding of existing materials and come up with new kinds of substances [167].

4. Quantum optimization: Qubits are utilised to find the cost functions in the optimization problems, which can be used to solve problems in finance, energy, and logistics [168].

5. Quantum sensing: Qubits can be used to measure various physical properties, such as magnetic fields and temperature, with higher precision than classical sensors [169].

4.3 Quantum entanglement

Quantum entanglement is a phenomenon in quantum mechanics where two or more quantum qubits become correlated in such a way that it is mathematically impossible to explain the condition of one qubit without simultaneously describing the state of the other, even when separated by large distances. The concept of quantum entanglement has been central to the development of quantum information theory and has many practical applications. Some examples, including quantum computing, quantum communication, and quantum cryptography. In particular, research in the field of quantum random number generators has shown the entangled photon pairs can be used to deliver actual random numbers, which are important to various cryptographic protocols [170]. These are just a few examples, but there are many other potential applications for quantum entanglement in classical networks. In entangled systems, the system's wave function cannot be factorized into the wave functions of the individual components. This implies that the states of the two components are connected to one another. Mathematically, this correlation is described by the Schmidt decomposition, which expresses the wave function of an entangled system as a product's sum of the states. The coefficients of these product states determine entanglement degree between the entangled bits. A system of two entangled qubits can be described by the following joint wave function [171]:

$$|\psi\rangle = \frac{1}{\sqrt{2}}(|0_A 0_B\rangle + |1_A 1_B\rangle) \tag{2}$$

where $|0_A 0_B\rangle$ and $|1_A 1_B\rangle$ are the basis states of the two qubits, i.e. A and B. The $|0\rangle$ and $|1\rangle$ represent their respective polarization states (horizontal or vertical, for example) A wave function for a system of three entangled photons can be written as:

$$|\psi\rangle_{abc} = \left[\frac{1}{\sqrt{3}}\left(|1\rangle_A|1\rangle_B|1\rangle_C + |1\rangle_A|0\rangle_B|0\rangle_C + |0\rangle_A|1\rangle_B|0\rangle_C\right)\right] \tag{3}$$

Where A, B, and C represent the three photons. However, the probability amplitude of measuring any single photon in the state $|1\rangle$, or $|0\rangle$ is $1\sqrt{3}$. This wave function represents an entangled state as each photon's state is correlated with the other photon's state, meaning that measuring the state of one photon will instantly affect the state of the other photons, seamlessly. In other words, the information can be transferred amongst the entangled photons with the speed faster than the speed of light. This phenomena can be utilized in the new coming generations to save time and power. As one classical bit drives the laser that pumps the BBO crystal, not only multi photons can be generated, rather, these photons own a hidden channel that also can be utilised. Subsequently, entanglement has the potential to be used in ORAN networks to reduce signaling costs. For instance, using entanglement to transmit and share information amongst the DUs using less amount of power compared to traditional network, which reduces the overall signaling costs. Additionally, entanglement can be used to provide secure communication in ORAN networks, which is important for protecting sensitive information. This can help to mitigate security risks and reduce the need for expensive security measures, which can also help to lower signaling costs. Based on the above advantages, this chapter proposes entanglement based method to reduce the cost of power and delay in ORAN networks by exploiting the use of hidden variable property. The latter manifest transferring the information amongst the entangled photons of the CUs and DUs.

4.3.1 Quantum cloud networks

A laser can be derived by the classical bits of a specific classical UE; the laser then pumps the nonlinear crystal, producing the entangled photons. These photons are transmitted to the RRHs where this UE resides, using an optical fiber or wireless channel. Subsequently, the photons are detected at the RRHs, each with a specific photon state, and the classical bits are recovered. As a result, this process has duplicated the classical bit to several bits at no additional expenses. When the UE travels to the neighbouring RRH, the information is served immediately using these redundant bits (already sent to the destination RRH at the time of photon generation). The need for an X2-AP framework protocol for handover signalling then is mitigated. The legacy problem of the cloud radio access network is that it allows the UE to connect to the cloud center so as its data to be processed, then these data are sent to the UE through its RRH. The network delay is consequently increased since the distances to the UEs are increased. Moreover, further delay will be caused by the control plane, mostly the handover process. If the handover takes place, multiple packets will be exchanged between UE, destination BBU, source BBU, serving and packet gateways, and mobility management units, this causes the cost of delay and power consumption to be at high levels [172]. Therefore, the proposed approach uses entangled photons as direct transmission signals between the associated handover units to reduce these costs. However, this study utilises the hidden channel between the interconnected photons, where changing the polarization state of one photon directly affects the others.

4.4 Quantum handover

The classical handover procedure can be described, as follows:

1. The UE receives a power level from a target RRHs and reports these to its existing RRH (source RRH), the UE uses RRC control signals for all possible target RRHs.

2. The target RRH is selected to be the based on which one the UE receives higher power from.

3. The source RRH sends a handover request to target RRH to plan the handover method with the required information (e.g., RRH detail, UE context, resource blocks mapping).

4. The target RRH shall track the availability of necessary resource, and sends a confirmation to the former RRH.

5. In the meantime, the UE will aim to access the target RRH, transmitting the message to its target RRH 'RRC Link Setup Complete.' The latter then sends to the MME a message telling the UE that its RRH has been updated.

6. The MME sends UE details and the current position to theSGWandPGW. Subsequently, theSGWsends downlink packets to the target RRH rather than the source RRH and recognizes the MME.

7. Finally, the target RRH calls on the source RRH to finally release the UE. This led to the end of the transition process.

It is worth mentioning that the handover process in the cloud architecture happens in the cloud center, where the source and target BBUs are all together in the same place. In contrast, the quantum handover happens amongst remote parties. Below is some of the features for the quantum method.

1. Let us assume BBU1 serving RRH1 and BBU2 serving RRH2. While the RRH1 UE transferred to RRH2, after all, it serving (BBU1) could still be used, like photons (converted into conventional bits at RRH) are rendered from one bit of UE information. Again, The UE data is then doubled and directed to RRH2, saving power and time in the pool.

2. This means the UE can be moved to target RRH2 and still be served by BBU1. This matter is very important as the target RRH is not always ready for the handover, not supported by X2-AP or does not own the required resources on time. However, the UE's requirement for the status transfer is not requested to provide additional control signals with the target BBU.

3. It provides free channels to transfer photons between RRHs and Cloud Centers using optical fibers.

4. The study has shown that the X2-AP protocol faces a significant loss in classical communications, which can be described as unreliable and scalable [173].

5. Classically, the X2-AP interface can be upgraded to the latest in both BBUs, which is tedious and costly [174].

6. Some BBUs have no X2-AP interface within the network architecture traditionally; the S1 protocol is a replacement in this case. Two BBUs carry out the handover along with the MME. In this case, the interconnection approach applies to an optimal relief of X2 and S1 to carry out the switch.

4.5 System evaluation

In more details, the classical handover can be described in Fig.4.1:

The quantum handover is relying on performing each of the steps of fig 4.1 but utilising the photon states of the entangled photons, where $|\psi 1|$, $|\psi 2|$, $|\psi 3|$ and $|\psi 4|$ are the final photon states (after detection) offour entangled photons, as shown in fig 4.2.

In Fig. 4.3, when the UE of RRH1 moves to the next RRH2, the cloud sends the UE data to all the surrounding RRHs of the UE, enabling copy-free of such data, thanks to the generation process of entangled photons. Meanwhile, if the sending eNodeB informs the MME about the handover, the former can utilise the hidden quantum channel to pass the information to the latter. Passing the information can simply be implemented by changing the polarisation of the former, the latter will change

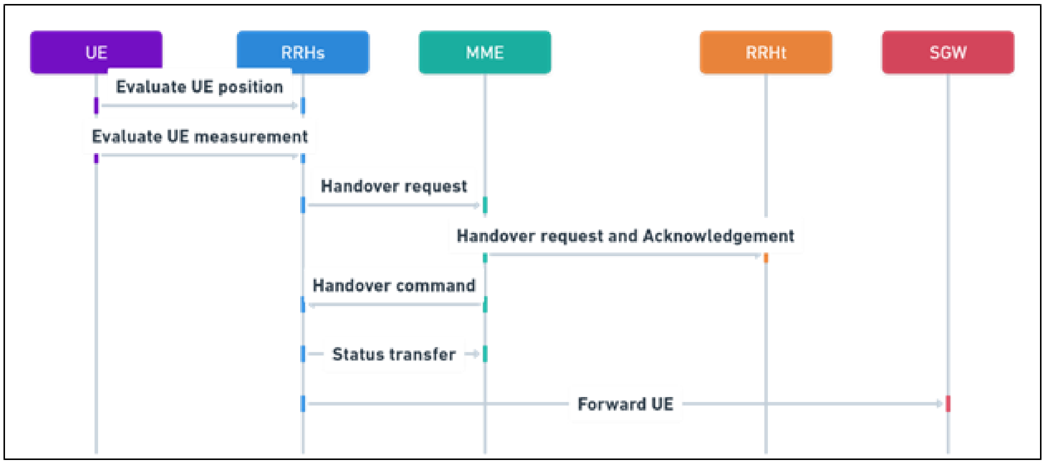

Fig. 4.1. Classical Handover.

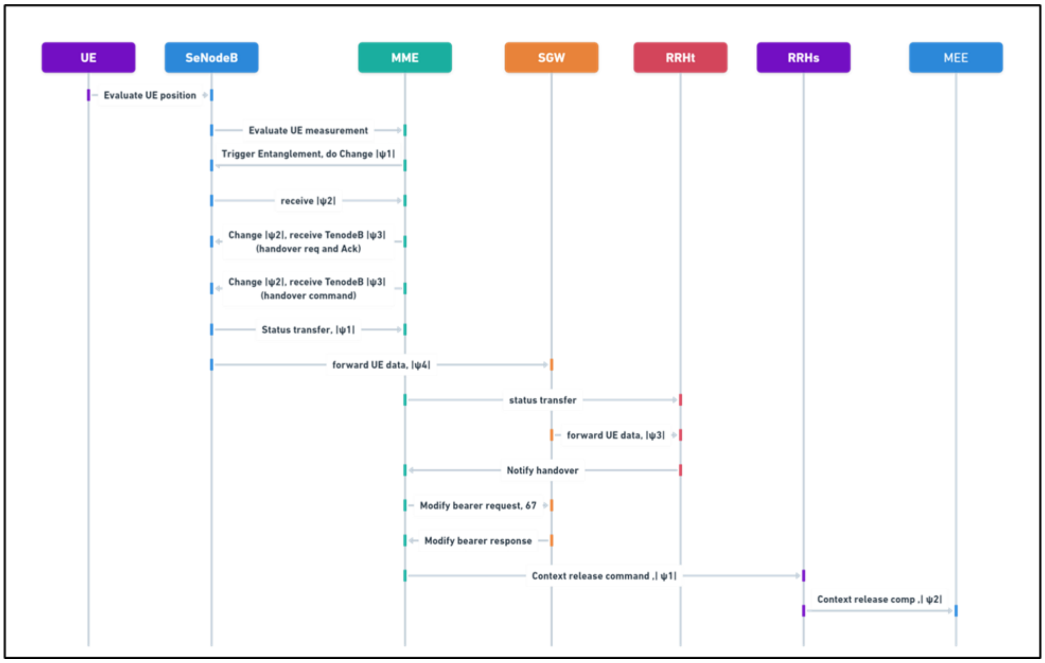

Fig. 4.2. Quantum Handover.

immediately at no time.

We first examined the UE position, where the UE informs the serving RRH of its RCC measurements. Once the decision is made, several connections has to be made

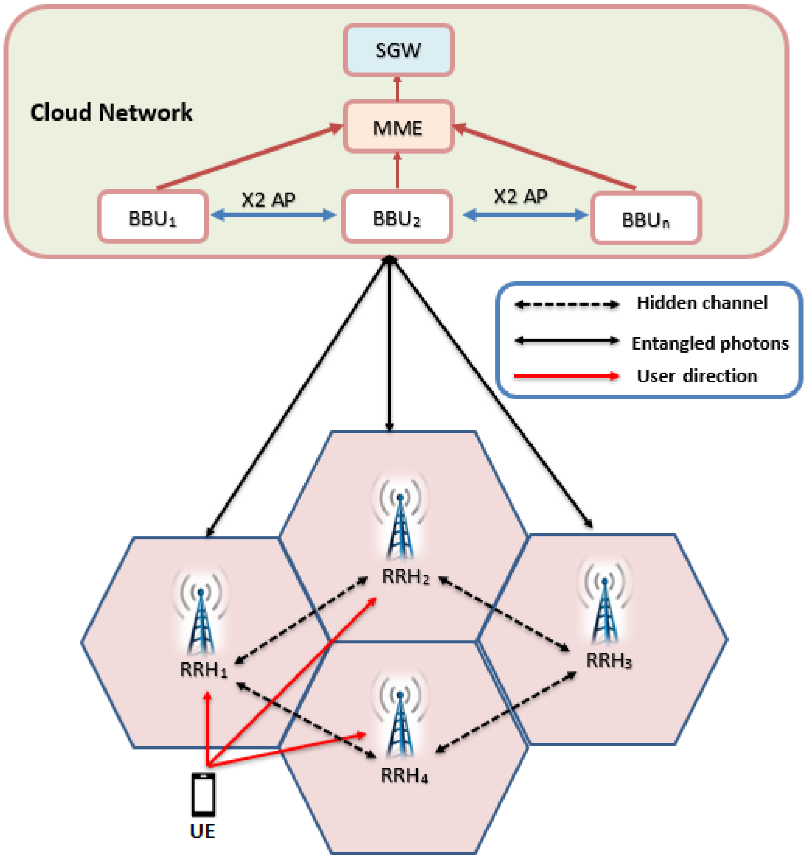

Fig. 4.3. Quantum handover process architecture.

to finally release the UE to the target RRH, as shown in fig 4.1. After receiving the measurement of the UE, the serving RRH sends communicates with the MME to inform about the handover process, the MME in turn, informs the target RRH and finds if it has the required resources, with handover request and acknowledgment signals. Then the MME commands the RRHs of the handover. The later sends the UE status to the MME and UE data to the SGW to establish the new channel for the UE. Then more communications to be done amongst the participants to finally release the UE. In the quantum handover, presented in the fig 4.2. In the latter, the classical signals are replaced with state changing procedure. The advantage of such method is the time reduction. The polarisation of the states, once it is perturbed, the other correlated states are all responded and be collapsed. This situation can happen amongst whatever units that participate in the handover procedure.

4.6 Classical delay

The time delay of this process can be analysed by evaluating the time of each sub-control operation due to the handover process. Although, the classical handover timing diagram depends on the latest technologies related to manufacturing the servers responsible for processing, manipulating, and sending the necessary control signals. However, the overall timing for the classical handover procedure is taking a remarkable cost that may cause of the outage in the UE's connection, loss of power, increased delay and lack of network reliability. The delay is analysed in many steps before knowing the differences between classical and quantum methods. We have denoted the MME with m, sending RRH with s, target RRH with t, and gateway with g. That is D_{sm} means the delay between the sending RRH and MME, D_{mt}, denote the delay between the MME and target RRH. Moreover, the delay between the MME and serving gateway is denoted by D_{mg}, and so on.

The overall delay of the classical method is the combination of processing delay and channel distance delay. The processing delay in the classical handover is known in the range of several milliseconds. If the handover request operation is evaluated, the delay of sending, channel and receiving will be evaluated, as follows:

$$D_{sm} = D_s^p + d_{sm} + D_m^p \tag{4.2}$$

where D_s^p represents the processing delay of the sending RRH, d_{sm} is the distance delay between the sending RRH and the MME, and D_m^p is the processing delay of the MME. Moreover, the delay of the handover request between the MME and target RRH (D_{mt}) is calculated as:

$$D_{mt} = D_m^p + d_{mt} + D_t^p \tag{4.3}$$

where D_t^p is the processing delay of the target RRH, and d_{mt} denotes the distance between them. This procedure will continue until the UE is finally released.

4.7 Quantum delay

In the quantum case, there also be a delay that is originated from the process of generating the entangled photons. The delay in the quantum case mostly happens in the circuit responsible for synchronising, elaborating and measuring the photons states amongst the different RRHs. In addition, there is another delay that happens when the tagged RRH informs other RRHs about its measuring state, classically, so the other RRHs detect whether their collapse states are correct or not. Accordingly, the RRHs error-correcting the received states, quantum wise.

The first delay consumer unit is the polarisation measurement at the receiving side, where the sending RRH measurs its polarisation due receiving a classical signal revealing the state of the sending RRH. Subsequently, the receiving RRH examined if its final state was correct or not. If not, correcting this state is mandatory using quantum error correction by re-sending the entangled photon. Suppose the classical signal being transmitted at the same time of measuring the state. We have denoted the delay due to the classical channel at each unit by d_c, this delay will be for all participating units. However, the delay of receiving the entangled photons is divided by two parts: the first is the delay of receiving detector at each unit, or called response time, denoted by D_{res}, second, the delay of translating this photon to a classical bit, denoted by D_{dri}^p. Hence, the total delay in the quantum case is summarised as follows:

$$D_{qd} = D_{res} + d_c + D_{dri}^p \tag{4.4}$$

4.8 Quantum consumption

We have assumed the total power consumption of the network, denoted as P_{QT} included two main parts: the traditional power consumption and the quantum. The power consumption of the traditional cloud is $P_{traditional}$, and the power consumption of quantum side is $P_{quantum}$, given by:

$$P_{QT} = P_{quantum} + P_{traditional} \tag{4.5}$$

The former mainly contains the BBUs and the RRHs. The BBUs are responsible for processing the base band signals and the arrived/transmitted packets of the UEs. The server power consumption is denoted as P_{server} , where a group of servers assemble the cloud center. There are other consumptions within the cloud such as the power overhead, and fiber losses. It is worth mentioning that the server consumption itself is not a fixed value, it is directly proportional to the number of processed packets, i.e. the bandwidth (BW). The change in its consumption ∂P_{server} to the change of the band-width ∂BW is equivalent to a constant, as follow:

$$\frac{\partial P_{server}}{\partial BW} = \alpha P_{server} \tag{4.6}$$

when solving this equation, it produces:

$$P_{server}(BW) = P_{server}^{initial} exp^{\aleph BW} \tag{4.7}$$

The server power consumption as a function of the bandwidth is the initial power consumption that is affected by the constant and the bandwidth. When there is no bandwidth (no load), the server power consumption is only its initial power consumption, i.e. idle mode of operation. In addition, we assume the total number of operating servers is C, and the total servers power consumption is represented by $P_{servers}^{T}$. However, the cloud, as mentioned earlier, included other consumptions, that are also proportional to the total servers consumptions. These losses are summed by, AC-DC, DC-DC, and cooling power consumptions. Generally, these consumptions are due to power losses. For example, the AC power is not efficiently converted to the DC power (required for operating the servers). As such, the DC power is not perfectly converted to the required value of DC power (required for each unit in the server and the cloud). Hence, we have assumed this consumptions as power losses. The AC-DC is represented by σ_{AC}, the DC-DC consumption is denoted by σ_{DC}, and cooling consumption is represented by the factor $\sigma_{cooling}$. Subsequently, the cloud power consumption is formulated as follows:

$$P_{traditional} = \frac{P^{T}_{servers}}{\sigma_{AC} \times \sigma_{DC} \times \sigma_{cooling}} \tag{4.8}$$

And this is valid for only one cloud. If there is more than one cloud, the above formula is repeated as many as the cloud centers.

The other part of the cloud consumption is the RRH, we have denoted this consumption as (P_{RRH}). The power consumption of this unit contains the radio unit (P_{RADIO}), power amplifier (P_{AMP}). This unit is also submitted to the overhead losses, but not the cooling, as its consumption is low and does not requires cooling.

$$P_{RRH} = \frac{P_{AMP} + P_{RADIO}}{(\sigma_{AC})(\sigma_{DC})} \tag{4.9}$$

where $P_{AMP} = P^{t}_{r,ue/\sigma_{pa}}$ is formulated as the transmitted signal to the UEs $P_{rrh,ue}$, to its efficiency ηAMP. Hence, the total power consumption of the traditional part is updated to the following, as pursues:

$$P_{traditional} = \frac{P^{T}_{servers}}{\sigma_{AC} \times \sigma_{DC} \times \sigma_{cooling}} + \frac{P_{AMP} + P_{RADIO}}{(\sigma_{AC})(\sigma_{DC})} \tag{4.10}$$

The quantum part of the network can also be divided into two parts, the first part is the quantum cloud part, denoted as P_{QC}. The second part is the quantum RRH part, denoted as P_{QR}. In the former, there are several components that are required to perform the necessary quantum computations. It is to be noted that the uplink communications is always classical, and the downlink is quantum. This required a laser in the cloud center to pump the BBU crystal that generates the entangled photons and send them to the RRHs. The uplink procedure can be done classically and no need for the detector in the cloud. At the RRH, it is required several units, a detector to receive the photon, a driving circuit, and a polarisation synchroniser. Hence, the power consumption of the quantum cloud P_{QC} is equivalent to $P_{QC} = P_{laser}$, while the P_{QR} can be given as follow:

$$P_{QR} = P_{det} + P_{driver} + P_{synch} \tag{4.11}$$

Hence, the quantum power consumption can be summed as

$$P_{quantum} = P_{QR} + P_{QC} \tag{4.12}$$

4.9 Energy efficiency

Among the different network metrics, energy efficiency is an important metric to evaluate, considering power consumption as a parameter. In this work, the energy efficiency gain is evaluated to show the importance of the proposed method. The EE can be defined as the transmitted data rate (bits/s or bps) to the power consumption (Watt). This means how much data rate is transmitted when consuming one Watt of power, i.e. (bps/W). As a matter of the fact, each classical protocol happens at different bandwidth than the data plane bandwidth, in this work, we have assumed the bandwidth as 10 MHz.

$$CRate = \sum_{M}^{m=1} BW log_2(1 + \frac{P^T_{c,m} H_m r_m}{AWGN + I_m}) \tag{4.13}$$

Where CRate denotes the classical data rate, M is the total number of RRHs, AWGN is the additive white Gaussian noise, $P^T_{c,m}$ is the transmitted power of the $m - th$ antenna, and H_m is the channel gain of the RRH m. The term $r_m = d_m^{-\aleph}$ represents the path loss, where d_m is the distance of the RRH m to the target RRH. Here, $\aleph$ is the path loss exponent, and I_m denotes the interference from other RRHs on the tagged channel m. Consequently, the EE formula can be derived for the classical network as follows:

$$EE_{traditional} = \frac{CRate}{P_{traditional}} \tag{4.14}$$

In the quantum case, the data rate is already embedded within the entanglement quantum channel that happens instantly without classical considerations. However, for the sake of comparison, the bandwidth of the laser can be considered as the required bandwidth for the quantum case, as follows:

$$QRate = \sum_{M}^{m=1} BWlog_2(1 + \frac{P_{q,m}^{T} Pr(m, M)}{Loss_m}) \tag{4.15}$$

where

$$Pr(m, M) = \int d\lambda \rho(\lambda) P_{a1}(m, \lambda) P_{an}(M, \lambda) \tag{4.16}$$

denotes the coincidence probability amongst the measurements of RRH m and other RRHs M. $P_{a1}(m, \lambda)$ is the detection probability of particle a in the direction of RRH m, sharing the same value of the hidden variable λ. Subsequently, $P_{an}(M, \lambda)$ is the detection probability of particle a_n in the direction of other RRHs M, where a_n is the indication of particle number n, and $n \in 1 : N$ denotes the total number of entangled photons. In addition, $\rho(\lambda)$ represents the probability of the produced photon state. $P_{T,q,m}$ represents the transmitted power of the laser of the RRH m. $Loss_m$ denotes the network's loss budget on the RRH m, which includes the number of fiber splices, connectors, dispersion, and distance. Subsequently, the energy efficiency (EE) can be calculated as:

$$EE_{quantum} = \frac{QRate}{P_{quantum}}) \tag{4.17}$$

4.10 System complexity

The complexity of the proposed methods relies upon the continuous moving UEs. The proposed method is aimed to surround the moving UE with entangled photons so as its data plane constantly be available and the hidden channels can operate. When the UE moves to cells that are not within the entanglement zone (where the first set of photons are distributed), this case causes the UE to shift to the classical handover. This problem can be realised by predicting the direction of the UE and providing the extra cells with the entangled photons. Another solution is to provide more entangled photons in all directions around the UE. Another problem is that the RRHs are practically not uniformly distributed, based on hexagonal or circular shapes. Our work used the Poison point process to deploy the RRHs, a more practical-oriented paradigm that conveys real-time cell shapes. This matter requires an optimisation process to predict which RRH is closer to the UE and represents its surrounding cell. However, the more entangled photons to be used, the more cells can participate in the UE perimeter. Non the less, this process must continue to operate as long as the UE moves, providing a collar coverage for the next direction of the UE. However, generating more entangled photons is more complex than fewer photons, so the states of the generated photons become more challenging to distinguish. This matter requires more caring on the receiving side so that the tagged state will be purified.

4.11 Results and Analysis

There are participant units involved in the handover process, these are source RRH, target RRH, MME and SGW, we have assumed the distance of the source RRH to the MME100 km, the distance of MME to target RRH is 100 km, the distance between source RRH to SGW is 100 km, while the distance of SGW to target RRH is 100 and finally, the distance of MME to SGW is 50 km. These five distances have been suggested to show the existing connection amongst these parties no matter how many repetitive connections happen during the handover process. These distances are used to produce the channel delay of these wireless links. This wireless link can easily be replaced with optical fiber channels to compare and show other results of this work. In addition, the processing delay of the source RRH is assumed to be 3 ms, the target RRH is 3 ms, the MME unit is 15 ms and SGW is 5 ms. These has been added to the processing delay to produce the final delay that is shown in Fig 4.4.

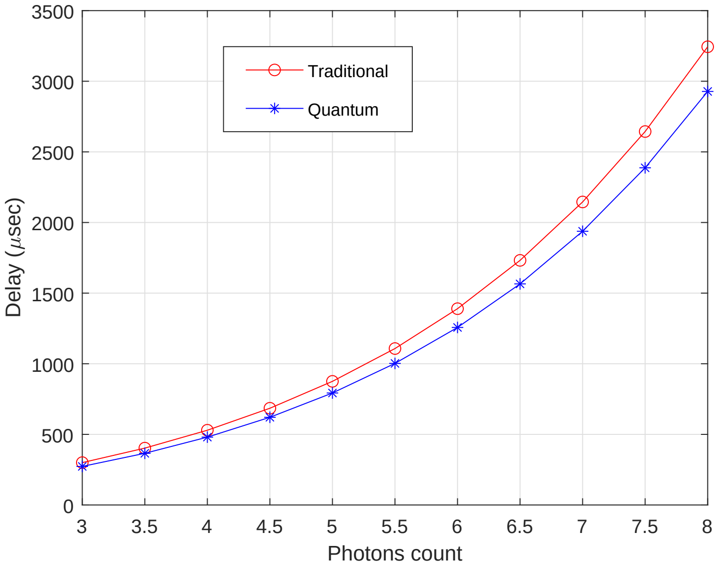

Fig. 4.4. Latency of the networks, quantum and traditional with respect to the number of entangled photons.

In the two-photon scenario, more delay will be produced than in the three or four photon cases due to more ping-pong signalling required. This means the more photons to be generated, the more efficient the system will perform. Note that when calculating the final delay of the source RRH-MME link, the processing delay in the source RRH occurs 5 times, so does the MME, as in fig 4.1. Hence, the total delay of this link is equivalent to the link delay, in addition to 5 times the processing delay. Similarly with other links, such as MME-target RRH shown in Fig 4.5.

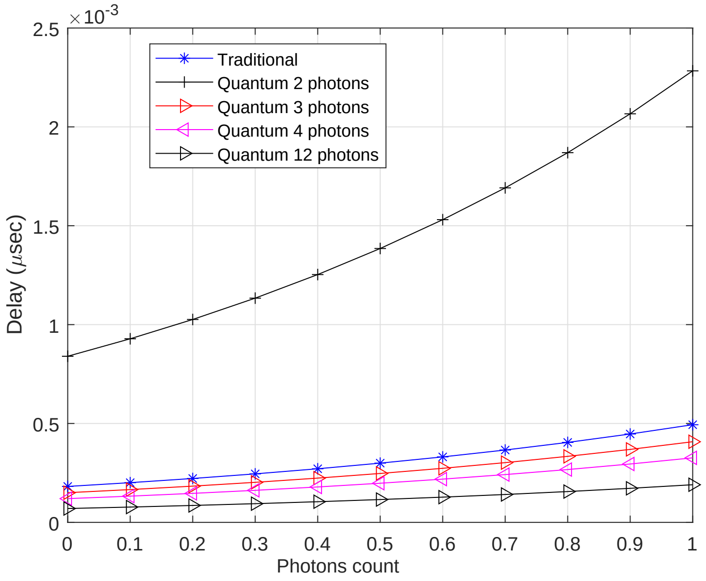

Fig. 4.5. Latency of the networks, quantum and traditional showing the effect of delay gain when using different number of photons.

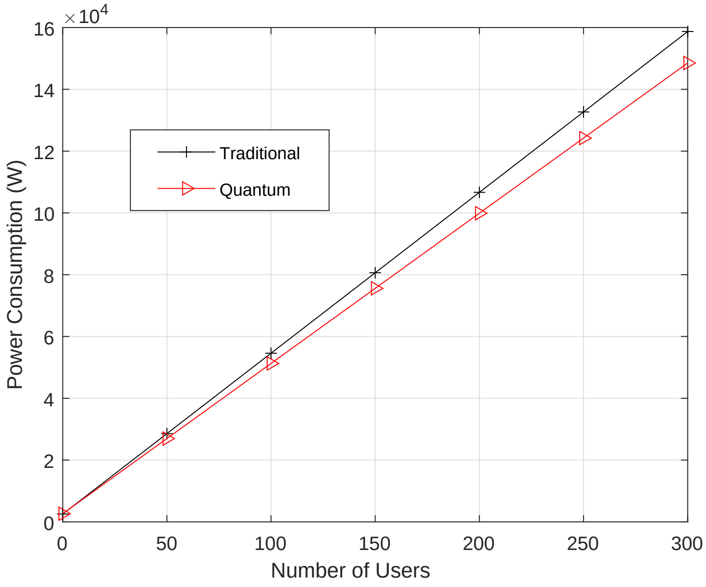

Fig. 4.6. Power consumption with respect to the number of UEs, when the X2-AP protocol consumes only 10% of the classical server power consumption.

Subsequently, the total delay of the traditional case is produced by jointly adding the delays of all links. The delay in the quantum case is also produced the same way, the processing delay of the laser, detector and the driving units, as shown in Table 2, have been jointly added to the total quantum delay.

In Fig 4.6, the power consumption has been presented with respect to the number of UEs. We have assumed the number of BBUs is 20, RRHs is 50. We also assumed the worst case scenario, where the X2 protocol consumes only 10% of the power consumption of the classical server, this amount has been deducted in the quantum case to gain such power. It shows when the number of UEs increases, the amount of power saving increases too. However, practically speaking, the network may contain thousands or millions of UEs that move constantly during the day. Hence, this saving can be further increased.

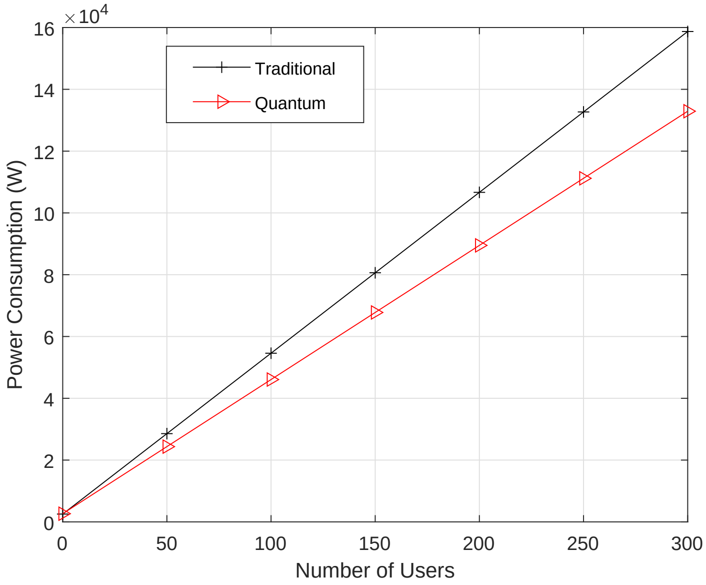

Fig. 4.7. Power consumption with respect to the number of UEs, when the X2-AP protocol consumes 20% of the classical server power consumption.

In addition, Fig 4.7 shows the power consumption of the two networks when the X2 protocol consumes 20% of the power consumption of the classical server. This case has gained more power as it reduces the amount of the classical X2 handover from the quantum case.

Later, the power consumption has been utilised to produce the energy efficiency, the average data rate was first calculated using the channel capacity formula. In the latter, the power from the RRH to the UEs was distributed based on the UEs distances to the tagged RRH, the nearest the UE to the RRH, the less received power. Additive white Gaussian noise has been suggested, the channel gain is also calculated.

Moreover, Fig. 4.8 shows the energy efficiency of the network when the power consumption is 10% less in the server compared to the classical one.

Where in the case of 20%, shown in Fig. 4.9, the energy efficiency of the quantum network will be further increased. However, the energy efficiency and the power consumption behave differently because in the former, the data rate will drive the

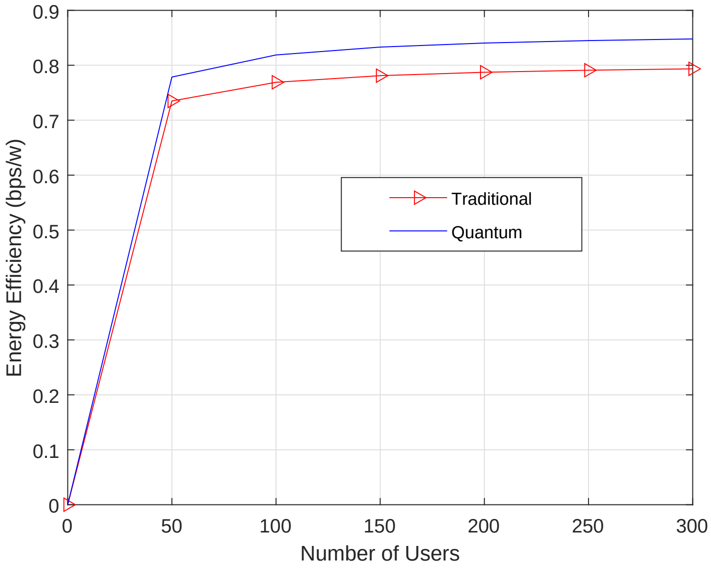

Fig. 4.8. Energy efficiency with respect to the number of UEs when the power consumption is 10 %.

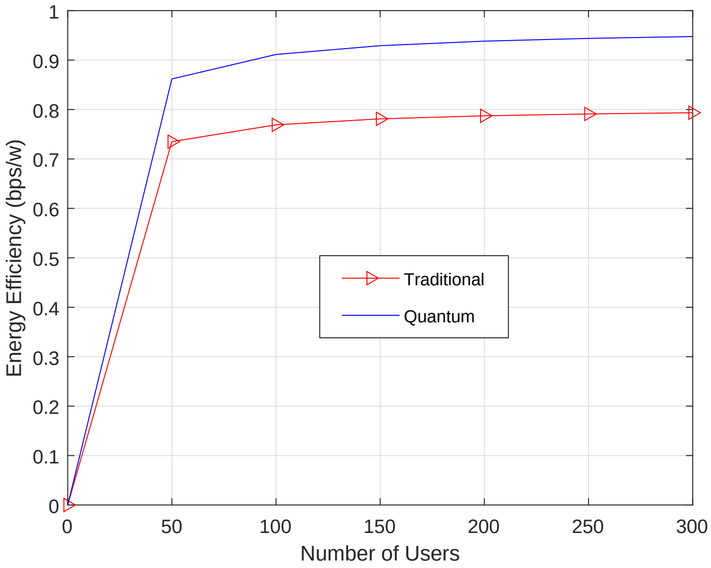

Fig. 4.9. Energy efficiency with respect to the number of UEs when the power consumption is 20 %.

Table 4.1

Model Parameters II

Factor	Value	Unit	Factor	Value	Unit
D_s^p	3	ms	d_{sm}	5	ms
D_m^p	15	ms	d_{mt}	5	ms
D_t^p	3	ms	D_{res}	1	-
d_c	1	-	D_{dri}^p	1	-
D_{dri}^p	1	-	$P_{servers}^T$	0.01	W
σ_{AC}	0.9	-	σ_{DC}	0.91	-
$\sigma_{cooling}$	0.92	-	P_{AMP}	29.7	W
P_{RADIO}	12.9	W	α_{AC}	0.8	-
α_{DC}	0.8	-	P_{det}	1	W
P_{driver}	1	W	P_{synch}	1	W
BW	10	MHz	H	1	-
$AWGN$	-10	dB/Hz	P_q^T	43	dBm
$Loss$	-3	dB			

increment of the power consumption towards exponential and linear behaviours, at the same time. First, exponential behaviour can happen as the UEs are still bandwidth and power hungry, which drives the average data rate to exponentially increase from zero to higher values while serving almost first 50 users in the network. After that, the scarce resources of the system urge to share the bandwidth and power transmitted amongst all the 300 UEs, which makes the system increase almost linearly while increasing the number of UE. It is worth mentioning that the number of UEs may fluctuate at each Monte-Carlo iteration as Poison point process distribution has been implemented to generate the UEs and the RRHs. Finally, the cloud centre has been assumed in the centre of the geographical area.

4.12 Conclusion and future work

This chapter showed how quantum entanglement can be used in classical cellular communications to improve the performance of the X2 application (X2-AP) protocol. We have concluded that the power consumption have been decreased to approximately 20% in the quantum case compared to the traditional network by increasing the number of UEs. Similarly, the delay has decreased while increasing the number of entangled photons used to connect the RRHs and other network parties. It is worth mentioning that the delay of two photons cases is more than the traditional case since there will be the enlarged number of background communications and synchronisation. However, by increasing the number of photons to four and more, the delay decreases compared to the traditional network by about 40%. Finally, the energy efficiency increases in the quantum case by decreasing the power consumption by about 10% as the number of UEs increases.

In the future, the quantum entanglement can be used not only amongst the RRHs, but amongst the RRHs and the cloud centre. This results in updating the cloud, MME and SGW without time cost because of the existence of a hidden channel. It was expected that this method can further improve quality of service regarding the time. However, the concurrent trade-offs have to be analysed regarding the power consumption and system complexity. The latter can be realized by the means of artificial intelligence and quantum computing algorithms to control the procedure of photon transmission, receiving, purifying the photon polarization's states, updating the handover participants and error correcting the undetected photons. Furthermore, increasing the performance of the proposed method to cover RRHs that are not connected to the same cloud center, this may impose additional complexity. The latter is represented by initiating more channels for synchronizing and tracking the UEs.

Chapter 5: Quantum-Enabled Method for Power, Delay and Energy Efficiency Enhancement in Open Radio Access Networks

ABSTRACT

ORAN adds new devices, protocols, standards, and reshapes the legacy design; which increases the PC and latency costs. These costs occur during the signaling process of many units in ORAN, such as the DU and CU. The purposes of the quantum mechanics based approach, specifically entanglement theory, and shows its impact on the traditional ORAN architecture's signaling. It prompts the current state of adoption with traditional networks. In addition, it models the PC, delay and EE for both traditional and quantum based ORAN. These systems are compared while considering different network parameters, such as the number of signaling messages and allocated bandwidth. The results showed that the quantum method has promised about 45%, 40% and 10% reductions in the EE, PC and delay, respectively when compared to the traditional ORAN. However, these reductions are affected by the number of transmitted messages and used entangled photons. If the latter is increased, more power and

time saving can be achieved. The application of the quantum entanglement concept can lead to reduced power consumption by reducing signalling overhead. This is based on the assumption that this technology is currently available, with a mathematical framework providing a model for behaviour.

5.1 Introduction

Since there are new functions and advanced units to be added to the traditional network, the signalling of the ORAN becomes under speculation. Signaling refers to the process of transmitting control information between different network elements to establish, maintain, and release connections. Signaling protocols are used to coordinate the various network functions and to ensure that the network is operating efficiently.

There are several types of signaling services and protocols that will be used in ORAN, including [34]:

1. Control and UE plane separation signaling: This protocol is used to separate the UE plane and control plane [175].

2. Session initiation protocol: For establishing, modifying, and terminating multimedia sessions such as Voice over IP (VoIP) calls.

3. Authentication signaling protocol: It is used to authenticate the UEs and to authorize accessing network resources.

4. GTP (GPRS tunneling protocol): It is used to manage the tunnels that are used to transport UE data.

5. S1AP (S1 application protocol): For controlling the S1 interface between the evolved packet core and evolved NodeB (eNodeB).

6. Open RAN radio interface protocol: It defines the interface between the DU and the RRU in the network, allowing for the use of different vendors' equipment [176].

7. Intelligent controller protocol: This protocol is used to control and manage the network, including the orchestration of different network functions and the management of resources.

8. Service-aware network controller protocol: This service is used to control and manage the network services based on QoS and security [177].

9. Fronthaul interface (O-FI) protocol: This protocol defines the connection between the CUs and DUs, allowing for the use of different vendors' equipment and software [34].

However, there are main higher level protocols amongst the different units that can be summarised as follows [34]:

1. X2, or Xn interface: This is the signalling interface between the traditional eNodBs or green NodeB (gNB) and other gNBs. It's purpose is to exchange management information amongst CUs [119]. Note that gNB is the term that is used in the next generation base station.

2. E2 interface: This is an interface for exchanging control and management information between the core network and the gNB, specifically the CU, DU, or even the other eNodeBs, with a focus on energy efficiency metrics.

3. E1 interface: This is a low-latency interface for exchanging control and management information between the CU-u (UE plane) and CU-c (control plane).

4. A1 interface: The orchestration platform makes use of this interface in order to connect the non-real time and near-real time units.

In ORAN, the DU can be virtualized and a single CU is connected to many DUs, each DU (or virtual DU) is then connected to the RU; that serves the UEs. Several protocols and connections are established between these two important units, i.e. the DU and CU. The system's performance suffers when the cost of processing new functions rises, leading to high energy consumption and longer response times. It is worth mentioning that for traditional LTE system, the latency of the control plane and the UE plane must be less than 100 ms and 10 ms, respectively. In addition, processing at the layer2 and Layer3 levels in the UE and eNodeB accounts for about 54% of the total latency in the control plane [178]. Although the cost of signalling is affected by different parameters, for instance, type of processors, number of UEs and type of service, yet, the cost of signalling is considerably high in the coming generations, see for example the X2 handover cost in [172], [179], [63]. The CU-DU can communicate to perform different functions, such as integrated access and backhaul, including UE and control planes, UE to network relay, intra-handover, inter handover, intra DU-CU centralized re-transmission, etc. On top of that, the virtualisation of the DUs or CUs within one physical server also increases the time required to process the packets, more than ever [62]. Hence, it is crucial to mitigate the signalling cost and establish unbeatable methods. The quantum mechanics has recently evolved as a unique solution to the legacy problems that came to a limit, those classical methods cannot solve. Specifically, the cost of signalling is non reducible because they shape how the network is designed, operated and offers seamless control. This is in contrast to physical layer, where the full advantage of optimisation techniques are exploited, including bandwidth and power resource allocations. Consequently, invincible problems require revolutionary solutions. Quantum capabilities are not fully used in classical communications due to the current limitations of quantum technology. Quantum communication systems require specialized equipment and infrastructure that is not widely available. Additionally, quantum systems are more complex and difficult to operate and maintain compared to classical systems. However, quantum technology has the potential to provide certain advantages in communications, such as increased security and faster data transfer speeds. For example, quantum key distribution (QKD) can provide secure communication via utilising the characteristics of quantum mechanics to ensure transmitting the information is secure and cannot be intercepted or tampered with. Hence, if well exploited, it can unleash the potential of classical communications [180] Moreover, the quantum entangled photons are utilized to replace the classical way of signalling amongst the different units in ORAN. This method makes use of the spare photons that are generated from one classical bit, and use those photons to save the power. In addition, the hidden channel amongst these photons is exploited to transfer the information amongst the DUs and CUs instantly, and with less cost.

5.1.1 Contributions

1. Since transferring the practical experience of the quantum domain to empower the classical networks is not fully realised, this work proposed an entanglement based method to mitigate the traditional signaling costs. This method was adapted to the mid-haul region amongst the CUs and DUs, where a large percentage of the network costs occur.

2. In ORAN, not only the classical burden is still active, but there are added protocols, units, and standardization that increase the PC of ORAN. Hence, a power model is proposed. In addition, a power comparison has been made for both traditional and quantum based ORAN.

3. The time delay in ORAN is also an important metric to be reduced. This work modelled and analysed the network time delays of the traditional ORAN with the quantum based method.

4. Based on the number of transmitted messages, a comparison of the EE is presented. Beforehand, the channel capacity for both quantum and traditional systems is analysed.

5.2 Quantum-Classical procedures

5.2.1 CU-DU procedure

Amongst the CUs and DUs, there are extensive signalling procedures that are fully occupying the network resources, including control and UE plane. At the same time, costing the network time and energy. For example, if the UE is establishing a connection, the procedure shown in Fig. 5.2 is activated to start the cell activation. However, the CU is responsible for transmitting the data to many DUs at the same time, composing a fully connected-like topology [181].

To clarify, the signaling costs associated with UE-initiated cell activations, detailed in Algorithm 1, represent just one component of the broader signaling expenditures in ORAN. This study comprehensively addresses all signaling costs between CUs and DUs, evaluating their impact on network efficiency and operational overhead. There are many signalling procedures between these two units.

Algorithm 1 UE cell initiating

```
RRC inactive transition procedure.
RRC to other RRC, inactive state procedure.
RRC connection reestablishment procedure.
Managing multiple tunnelling for F1-C.
UE initial access procedure.
Bearer context setup over F1-U.
```

5.2.2 Quantum Signaling

Traditionally, the number of messages that are transmitted amongst the different units depends on the number of connected UEs, type of signalling, and number of signalling messages. However, one example of traditional CU-DU transmitting/receiving

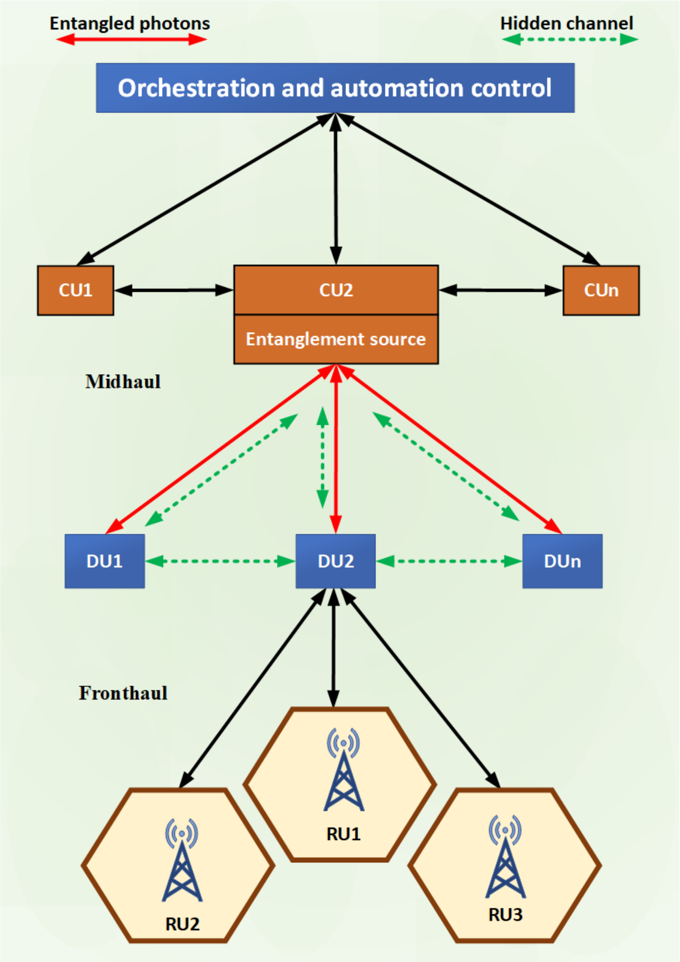

Fig. 5.1. System architecture of entanglement based open radio access network.

signals is shown in Algorithm (2), where radio resource control (RRC) signal is used to resume RRC transmission.

Keeping in mind this procedure only considers one DU, as per suggested ORAN, one CU is connected to multiple DUs. Hence, the cost of power and delay can be multiplied. Subsequently, the proposed procedure is shown in the Algorithm (3).

In Algorithm (3), the first photon ψ_{k1} is kept at the sender side (CU), while the second photon ψ_{k2} is sent to the receiving side (DU), the third photon (not shown in the Algorithm), is sent to other DUs with ψ_{k3}, $\psi_{k4}...\psi_{kn}$, where the total number

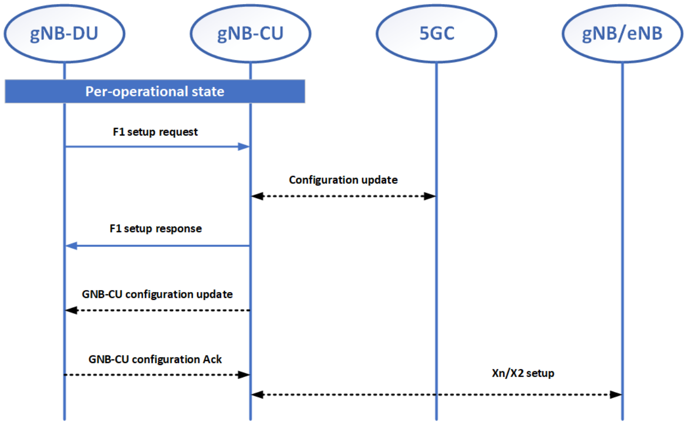

Fig. 5.2. Start up and cell activation

Algorithm 2 Classical: RRC state transition, Inactive

```
  CU-DU (Paging)
  DU-UE (Paging), UE-DU (RRC resume report)
  DU-CU (RRC message transfer)
  CU-DU (Context setup request)
  DU-CU (RRC message response)
  DU<->UE (RRC resume)
```

Algorithm 3 Quantum: RRC state transition, Inactive

```
  CU $\psi_{k1}$-DU $\psi_{k2}$ (Paging)
  DU-UE (Paging), UE-DU (RRC resume report)
  DU $\psi_{k2}$-CU $\psi_{k1}$ (RRC message transfer)
  CU $\psi_{k1}$-DU $\psi_{k1}$ (Context setup request)
  DU-CU (RRC message response)
  CU<->UE (RRC resume)
```

of entangled photons is N. Then in (DU-CU (RRC message transfer)), there is no need to install another entanglement source to transmit the photons and repeat the same process as the CU strategy. It is sufficient to alter the polarisation state of the received photon at the DU with ψ_{k2}, and the hidden channel is responsible for changing the polarisation state of the receiving CU, with ψ_{k1}. At the same time, DU ψ_{k2} may inform other DUs, with ψ_{k3} or more, about its declaration. Next, at the (Context setup request), the CU repeats transmitting its photons and informs other DUs about its photon state, requesting them to reveal theirs. This style continues until the last step of the protocol ends, i.e, (RRC message response). Nonetheless, this procedure is only an example of the quantum method. Amongst the CUs and DUs, there are tens of protocols and messaging. Whenever a different protocol is activated, the setup is kept the same and the difference will be in the packet construction.

5.3 System model

5.3.1 Power consumption

To calculate the PC of the cloud, we need to take into account the following factors:

- Server hardware: The PC of the physical servers and storage devices used in the cloud.
- Data center infrastructure: The PC of the data center facilities, including cooling, power distribution, and backup systems
- Network infrastructure: The PC of the network switches, routers, and other networking equipment used in the cloud.
- Software and applications: The PC of the software and applications running on the cloud servers.

However, it is crucial to evaluate the PC of the signalling as they contribute greatly to the dynamic PC of the system. If we assume the power of the CU's RB is b_{rb}^{cu}, the power of the DU's RB is b_{rb}^{du}, the number of RBs that are contained in the BW of CU and DU are RB_{cu} and RB_{du} at each time slot, respectively. In addition, the transmitted power of the CU and DU are P^t_{cu} and P^t_{du}, respectively. Hence, the power of the RB at each CU is equal to:

$$P^{cu}_{rb} = \frac{P^t_{cu}}{RB_{cu}} \tag{5.1}$$

Similarly for the DU's RB; is equal to:

$$P^{du}_{rb} = \frac{P^t_{du}}{RB_{du}} \tag{5.2}$$

This power is only consumed by one RB at a time slot, at a known amount of time when the transmission is continuous, this value is multiplied with as many as the number of time slots. There is another way to evaluate such power; is by calculating the number of transmitted messages. In CU and DU, the power transmitted for each message is given by:

$$P^{cu}_{msg} = P^{cu}_{rb} \times RB^{cu}_{msg} \tag{5.3}$$

And for the DU:

$$P^{du}_{msg} = P^{du}_{rb} \times RB^{du}_{msg} \tag{5.4}$$

Where RB^{cu}_{msg} and RB^{du}_{msg} denote the number of RBs in the message of CU and DU, separately. These are evaluated using the following formulas:

$$RB^{cu}_{msg} = \frac{Bit_{rb}}{Bit_{msg}} \tag{5.5}$$

$$RB^{du}_{msg} = \frac{Bit^{cu}_{rb}}{Bit^{ci}_{msg}} \tag{5.6}$$

Where Bit^{cu}_{rb}, Bit^{du}_{rb} and Bit^{cu}_{msg}, Bit^{du}_{msg}, denote the number of bits that are contained in the RB of CU and DU, and the sent bit in each message of CU and DU, respectively. The former can be obtained by:

$$Bit^{cu}_{rb} = Symbolts^{cu}_{rb} \times Subcarriers^{cd}_{rb} \times Bit^{cu}_{RE} \tag{5.7}$$

Where $Symblts^{cu}_{rb}$ denotes the number of symbols of the CU's RB, $Subcarriers^{cu}_{rb}$ is the number of sub-carriers per CU's RB, and Bit^{cu}_{RE} represents the number of bits in each resource element in the CU, which is based on the type of modulation, for example, in QPSK, the $Bit^{cu}_{RE} = 2$. The same style holds true for the DU, where:

$$Bit^{du}_{rb} = Symbolts^{du}_{rb} \times Subcarriers^{du}_{rb} \times Bit^{du}_{RE} \tag{5.8}$$

$Symblts^{du}_{rb}$ is the number of symbols of the DU's RB, $Subcarriers^{du}_{rb}$ is the number of sub-carriers per DU's RB, while Bit^{cu}_{RE} represents the number of bits in each resource element in the DU.

Now, it is possible to obtain the PC of the signalling process at the CU and DU, noted as P^{cu}_{sing} and P^{du}_{sing}, respectively, where P^{cu}_{sing}=$P^{cu}_{msg} \times N^{cu}_{msg}$ and P^{du}_{sing}=$P^{du}_{msg} \times N^{du}_{msg}$. Where N^{cu}_{msg} and N^{du}_{msg} denote the number of messages of the CU and the DU, respectively. However, this was only the dynamic load evaluation, the power supply for this signalling is evaluated as follows:

$$P^{cu}_{suply} = PC_{cu} * \frac{RB_{msg}}{RB_{cu}} + \frac{P^{cu}_{msg}}{\zeta} + P_{RF}\frac{RB_{msg}}{RB_{cu}} \tag{5.9}$$

PC_{cu} denotes the PC of the CU server, the PC^{cu}_{msg}/ζ terms represents the power amplifier's PC, where ζ denotes the amplifier efficiency. Subsequently, the initial PC of the CU and RF units are added to the above formula, symbolized as PC^{init}_{cu} and PC^{init}_{RF}, respectively. If we assume $PC_{initial}$= PC^{init}_{cu}+PC^{init}_{RF}, then:

$$PC^{cu,ini}_{suply} = PC_{initial} + PC^{cu}_{suply} \tag{5.10}$$

Finally, the PC of number of messages is given by

$$PC^{du,ini}_{suply} = PC_{initial} + PC^{du}_{suply} \tag{5.11}$$

The quantum method's PC $PC_{Quantum}$ can be calculated using the traditional one, but with some modifications. First, the quantum devices' PC is added to the quantum case. Subsequently, the amount of classical overhead, i.e. $(RB_{msg}/RB_{cu}), or(RB_{msg}/RB_{du})$ is removed from the corresponding CU or DU unit.

$$PC_{Quantum} = H(\psi)(P - initial - P_{suply}^{cu}) + P_{cost} \tag{5.12}$$

Where $H(\psi)$ is the hidden variable representative, Pcost is the PC of the quantum devices, which is equivalent to:

$$P_{cost} = P_{Laser} + P_{Det} + P_{Driver} \tag{5.13}$$

The P_{Laser}, P_{Det}, P_{Driver} represent the PC of the laser, detector and driver, respectively.

5.3.2 Quantum latency

It was assumed that the signal message has a duration of $\tau_{message}^{cu}$ message and $\tau_{message}^{du}$ second, for the CU and DU, respectively. A number of messages Ncu_{msg}^{cu} during a time interval called Δ_t^{cu} and Δ_t^{du} for the CU and DU, separately. Δ_t^{cu} can be written as

$$\Delta_t^{cu} = N_{msg}^{cu} \times \tau_{message}^{cu} \tag{5.14}$$

and

$$\Delta_t^{du} = N_{msg}^{du} \times \tau_{message}^{du} \tag{5.15}$$

The instant transmission of the information amongst the CUs and DUs implies that the network is no longer time restricted. The hidden variable contributes to the transfer of the state of the entangled states with zero time, while some time occurs due to the driving circuit of the receivers and the detector's response time of the quantum case. The total delay of the traditional method is modeled:

$$\tau_{traditional}^{cu} = \Delta_t^{cu,du} + \tau_{ch} \tag{5.16}$$

During the signalling, the τ_{ch} can be neglected in the quantum signalling, the total delay of the quantum method is modelled:

$$\tau_{quantum} = \tau_t^{cu,du} + \tau_{driver} + \tau_{det} \tag{5.17}$$

Where τ_{driver} and τ_{det} denote the time delay of the driver and detector, respectively.

5.3.3 Quantum capacity

The actual bandwidth of a quantum channel depends on various factors, such as the physical implementation of the channel, the type of quantum state that is utilised to encode the information, and the communication protocol [182]. The bandwidth in the classical and quantum systems is different. In the former, it refers to the amount of transmitted data per unit time. In the latter, it refers to the amount of quantum data that can be transferred in a certain amount of time, where the relationship between the bandwidth and qubit transmitted per second is linear.

Since it is possible to generate the entangled photons from discrete time pulsed laser with high percentage of fidelity [183], [184], it is possible to control the rate at which the photons are generated. In this section, the capacity of the quantum system is divided into three distinct categories: 1- The quantum bit capacity that can be directly under- stood by the number of quantum bits to be transmitted in one second. This capacity is based up on the bandwidth and number of generated qubits. As far as the frequency of the generated photons is considered. This frequency is half the frequency, double the wavelength of the pumping lasers frequency, which is still higher than classical frequency. Subsequently, the bandwidth is higher in the quantum than the classical. The quantum capacity Cq can be formulated as:

$$C_q = B_q log_2(1 + \frac{P_q}{N_q}) \tag{5.18}$$

where B_q, P_q are the bandwidth and power of the qubit. It is worth noting that the qubit bandwidth is much higher of the classical counterpart. In addition, N_q denotes the quantum noise effect, which can be formulated as the joint addition of the noise effects:

$$N_q = D(t) + E(j) \tag{5.19}$$

In this work, only the dephasing and amplitude damping noises are considered. Subsequently, $D(t)$ is the dephasing noise, it is given by:

$$D(t) = \upsilon_i \left[e^{-Yt/2}\right] |0\rangle \langle 0| + \left(1 - e^{-Yt/2}\right) |1\rangle \langle 1| \tag{5.20}$$

where Y is the dephasing rate, t is the time, and $|0>$ and $|1>$ are the basis states of the qubit, and υ is the effect of the entangled photon i. $E(j)$, $j \in 0$, 1 denotes the amplitude damping noise and can be described mathematically using the Kraus operator, as follows:

$$E(0) = \epsilon_i |0\rangle\langle 0| \sqrt{p} |1\rangle\langle 0| \tag{5.21}$$

$$E(1) = \epsilon_i [\sqrt{1 - P} |1\rangle\langle 1|] \tag{5.22}$$

where p is the damping probability.

2- The capacity of the hidden channel that shapes the number of transmitted information of entangled photons, and can be calculated using the following equation:

$$C_H = \max_{\rho_{AB}} \quad I(A : B) \tag{5.23}$$

where C_h is the hidden channel capacity, ρ_{AB} is the density matrix of the entangled photons, and I(A : B) is the mutual information between photons A and B. The mutual information is given by:

$$I(A : B) = S(\rho_A) + S(\rho_B) - S(\rho_{AB}) \tag{5.24}$$

where $S(\rho)$ is the von-Neumann entropy of the density matrix ρ, defined as:

$$S(\rho) = -\operatorname{Tr}(\rho \log_2 \rho) \qquad (28) \tag{5.25}$$

The channel capacity can be further simplified for specific entangled states, such as maximally entangled states. For a maximally entangled state, the density matrix is given by:

$$\rho_{AB} = \frac{1}{d} \sum_{i=1}^{d} |i\rangle_A |i\rangle_B \tag{5.26}$$

where d is the dimension of the entangled space. In this case, the channel capacity is given by:

$$C = \log_2 d \tag{5.27}$$

which means that the channel can transmit d bits of information per user terminal. 3- Quantum repetition Rate: Although the quantum band-width can be very high, it cannot be utilized unless high classical repetition rate is achieved. The classical repetition rate can affect the rate at which the quantum bits are generated. The process of generating the quantum bits is based on a nonlinear effect that happens in the BBO crystal, which is based on the high energy of the laser pulses [185]. These pulses are previously derived using classical bit rates. Hence, it is possible to write:

$$C_c = X \cdot f_{rep} \tag{5.28}$$

where C_c denotes the classical capacity, and f_{rep} is the repetition rate of the laser pulses. The non successful generation process is taken into account with the possibility of X <= 1. f_{rep} is responsible for generated the entangled photons's capacity C_q; with another probability, denoted as Y. As long as the entangled photons are generated, the hidden channel with capacity C_h is perfectly existed amongst the photons.

$$f_{rep} = C_h Y C_q \tag{5.29}$$

This means we can write the classical capacity in terms of quantum, as follows:

$$C_c = C_h[XYC_q] \tag{5.30}$$

Practically, this expression not only means two types of information are obtained from single one, in addition, the right term can be multiplied by the number of generated photons. In terms of EE, now it can be easily calculated for the traditional case using the following formula

$$EE_c = \frac{C_c}{PC^{du,ini}_{suply}} \tag{5.31}$$

and for the quantum case

$$EE_q = \frac{(C_c)(C_q)}{PC^{quantum}} \tag{5.32}$$

5.4 System complexity

- When the transmitting/receiving occurs amongst the CUs and DUs. Some photons may be spared. In this case, these photons can be exploited in other protocols that require communications amongst the CUs and DUs to perform different functions or different protocol.

- Although the hidden channel helps in transferring the information seamlessly, the holders of the entangled photons must declare the polarization state they have detected so that all participants are aware whether their state is correctly detected or not. Revealing this requires a classical channel amongst the participants, which is power and time costly. Hence, it is beneficial to evaluate this cost and consider it within the presented outcomes [186].

- The detection of the entangled state requires critical evaluation of the noise types that affect the entangled states. There are different noises' behaviours in the quantum channels that are different from the classical counterparts. Some common types of noise in the quantum channels are:

1. Decoherence noise: This type of noise results in a loss of coherence in the quantum state, causing the relative phases between different basis states to become random and unpredictable [187].

2. Amplitude damping noise: This type of noise causes the amplitude of the quantum state to decrease over time, leading to a reduction in the overall strength of the quantum state [188].

3. Bit flip noise: This type of noise results in a change in the value of the quantum bit, flipping a 0 to a 1 or vice versa [189].

4. Phase flip noise: This type of noise results in a change in the quantum state's phase, causing the sign of the state vector to flip [190].

5. Thermal noise: This type of noise results from the random motion of particles due to the heat energy in the environment, causing random fluctuations in the quantum state [191].

   Subsequently, understanding and mitigating the effects of noise is crucial for designing and implementing reliable quantum communication systems.

5.5 Results and discussion

The PC method has been compared with the traditional network from [57] and [55]. The different behaviours of the two systems require manipulating some of the parameters to obtain fair results. For instance, the number of RBs in each bandwidth might be different in reality as the bandwidth of the quantum system results from the type of generation process itself, as shown in [192]. Once translated to RBs, the construction of the produced results might be different. None the less, since the targeted parameter of comparison is the PC of the two systems, the bandwidth has been assumed equal for the quantum and traditional. Fig. 5.3 shows the PC comparison of the two systems, presenting the supply power with regards to the number of transmitted messages.

It is clear that the more transmitted messages for signalling, the more PC of the traditional network compared to the proposed method. This result was obtained when using only 6 entangled photons, 1 CU, and 5 DUs, ζ =32%. It was expected that the PC saving increases while the amount of photons, CUs, and DUs are increased too. The exponential shape of the traditional PC is previously verified in [62] that shows the dynamic load is exponentially increasing the initial PC of the network's unit. While in quantum PC, this behaviour is linear due to less affected by the dynamic load. However, the quantum PC scores more initial PC.

In addition of the P_{laser} = 1W, P_{det}=2W and $P_{drivers}$=2W. In the following Fig. 5.4, the supply power was compared for the two systems in regards to the bandwidth.

The decreasing behaviour comes from the assumption that the number of RBs of each message RBmsg is less than the number of RBs available in the CU RBcu. This means when increasing the bandwidth, the model tends to decrease in value. None the less, this depends on the type of transmitted message and the type of protocol. Overall, the quantum case showed less PC than the traditional and decreased the PC to about 40%.

In terms of delay, the distance between each DU and CU has been assumed equal to 20 Km, while the distance amongst the DU-DU is 5Km. In case of the DUs are virtualised in one server, the distance amongst the DUs is ignored as they reside within the same place. However, if the number of DUs is large and some of them are virtualised, the distance amongst the virtualised servers is considered. Fig. 5.5

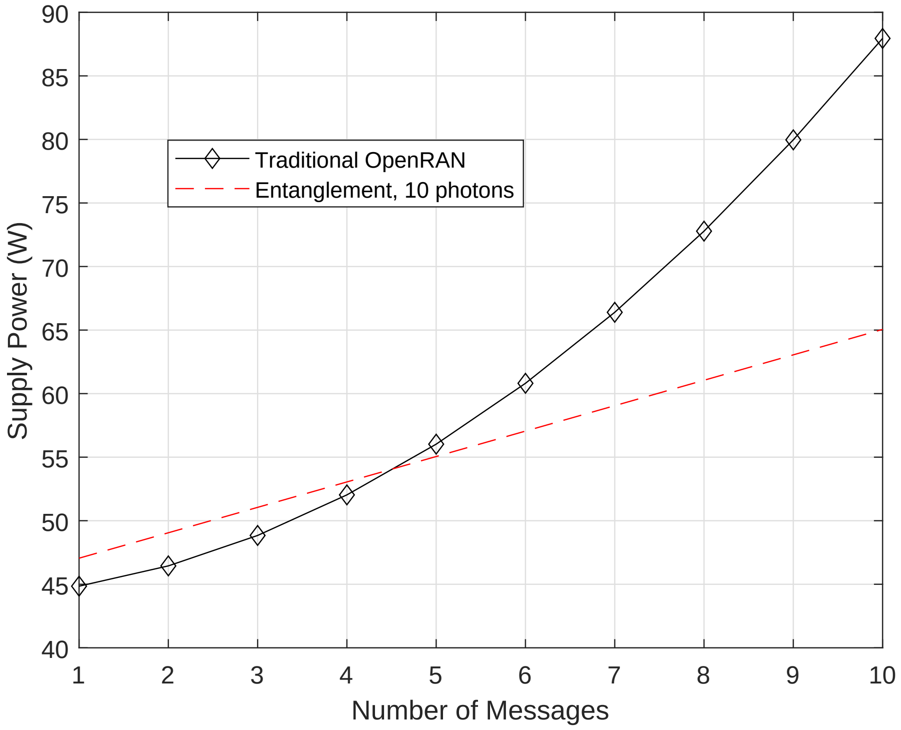

Fig. 5.3. omparison the PC of traditional and quantum ORAN with regards to the amount of messages.

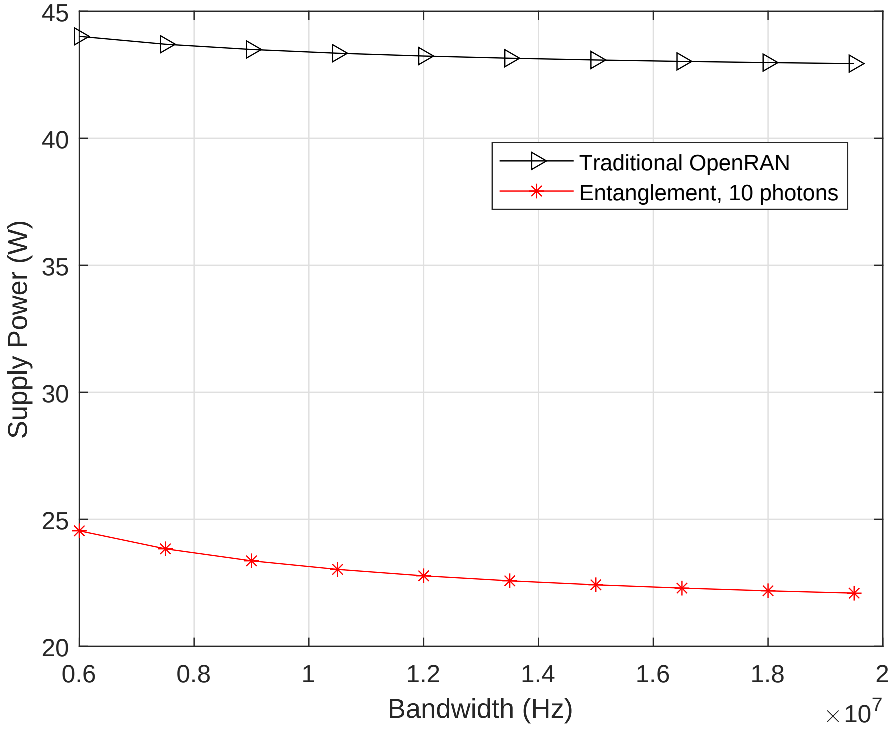

Fig. 5.4. PC comparison of the quantum and traditional ORAN with different bandwidth.

shows the delay comparison of the traditional and quantum ORAN with different numbers of messages. It was assumed the τ_{driver} and τ_{det} are both equivalent to 1 µsec. Subsequently, the quantum method has participated in reducing the time of signalling, including both the time period of the message and the channel time required for each message. Subsequently, more photons means more time reduction. Although the presented delay model has given clear indication of the channel latency behaviour, there still more analysis to be implemented within the ORAN networks. This delay might involve the CU and DU processing times, the ORAN controller processing and the virtualisation delay of the DUs.Moreover, the channel delays amongst these units contribute significantly to the overall end-to-end latency. By doing so, a clear evaluation of the total latency in the network will be envisioned. However, this work is limited to the proposed data as it is directly affected by the number of photons. Another factor that effect the results is when considering multiple cloud scenario. The signaling costs amongst the multiple clouds can also be further mitigated using multiple entangled sources.

Finally, Fig. 5.6 shows the energy efficiency comparison of the quantum and traditional systems considering the number of messages. In the quantum case, the bandwidth and amount of information transmitted via the hidden channel are greater than in the traditional case, leading to enhanced data handling capabilities. This has produced more bit rate in the quantum case, and subsequently, larger EE. The EE is calculated using the analysis in Section V-C to produce the channel capacity of both systems, then the capacity is divided by the PC of both systems. It is notable that the more control messages are transmitted, more EE can be achieved.

5.6 Conclusions and future trends

Entanglement based PC and delay reduction technique is proposed. This work showed that the potential for quantum networks is far more energy-efficient than traditional networks. However, its actual PC comparison depends on various factors

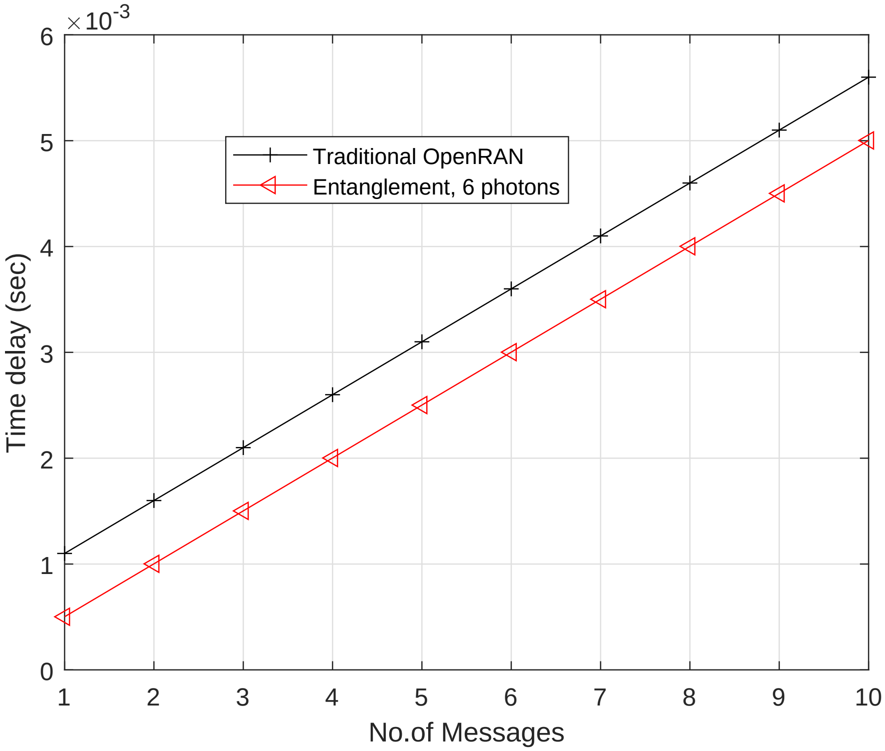

Fig. 5.5. Delay comparison of the quantum and traditional ORAN with different number of messages.

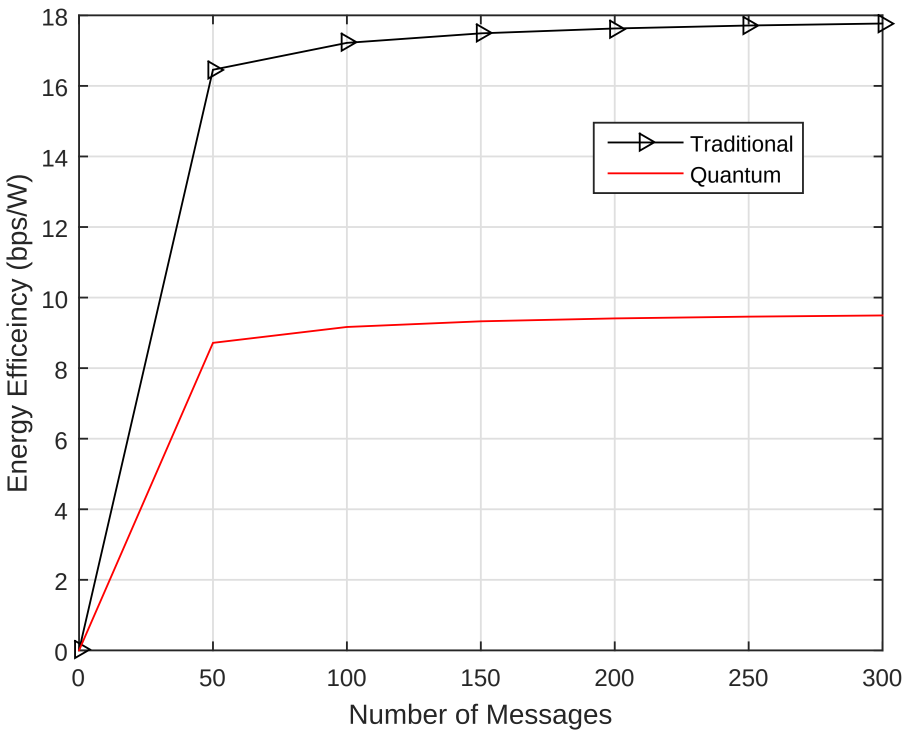

Fig. 5.6. Energy efficiency comparison of the traditional and quantum systems.

such as the specific implementation, network size, and desired functionality. This work also presented the PC delay and capacity models, by which, a comparison has been made between the traditional and quantum based ORAN. These metrics have been evaluated in correspondence to different parameters, such as the number of entangled photons, number of transmitted messages, system bandwidth and number of CUs and DUs. The model evaluated the metrics when the case of signaling among the CUs and DUs is considered. Subsequently, the model has produced noticeable amount of PC and delay savings, 40% and 10%, respectively. In addition, the EE has been enhanced by about 45%. Nevertheless, these percentages are subjected to the above mentioned network parameters.
It is important to note that the future dimension of the quantum domain is unlimited with new applications and adaptations to the classical field. Here we mention two interesting theories that can be applied within ORAN architecture, quantum security and quantum teleportation. One example of a hidden channel in entanglement is quantum teleportation. It permits the transmission of any chosen quantum state between any two particles via using entanglement and classical communication. This protocol can be utilized among CU, DU and RU, where heavy signaling cost occurs. In which, it is possible to transfer the quantum information through classical communication while minimal PC and time.

Quantum mechanics can further be used to enhance the security of ORAN in various ways to ensure its full potential is realized, including:

1- QKD can provide a secure key exchange between network nodes, ensuring the confidentiality of communication in the network.

2- Quantum random number generation can be used to generate secure random keys for encryption, providing an extra layer of security to protect against attacks such as brute force attacks.

3- Quantum cryptography can be used to ensure the authenticity and integrity of data transmitted over the network, making it more difficult for malicious actors to tamper with or steal sensitive information. Quantum technologies can help to address some of the legacy barriers associated with ORAN, but they also bring their own set of challenges, such as the need for specialized hardware and expertise, as well as the difficulties in integrating these technologies with existing networks.

Chapter 6: Optimizing the Energy Efficiency Using Quantum Based Load Balancing in Open Radio Access Networks

ABSTRACT

This chapter presents a novel approach to address the challenges in ORAN by including load balancing strategies and leveraging concepts from quantum physics, particularly by employing entanglement theory in classical communication systems. The proposed methodology facilitates the production of several quantum information units from a single classical information unit, with the objective of conserving the energy. Subsequently, this paper introduces a concise PC model for the ORAN architecture that is simplified, yet descriptive when compared to other models. The model effectively captures the fluctuations in traffic that servers handle and provides a comprehensive characterization of power usage within a virtualized system.

An optimisation problem is formulated with the objective of selecting ORAN servers for the quantum load balancing that are less energy-efficient, hence maximising user benefits. The resulted problem has been identified as a nonlinear problem (NLPP) with inequality and equality constraints. The utilisation of the Lagrange multiplier method is necessary when dealing with an objective function that exhibits non-linearity, as it provides a suitable mathematical framework. The numerical

problem-solving method 'fmincon' is utilised to compare two strategies, namely sequential quadratic programming (SQP) and the active-set approach. These strategies are specifically built to tackle nonlinear objective functions with constraints, however, with different convergence criteria. The aim of this study is to perform a comparative examination of energy efficiency (EE) between the CU servers and ascertain the server that exhibits lower EE, specifically for the purpose of load balancing that uses quantum theory. This research provides insights into the strategies that can be employed to achieve power efficient ORAN implementations. The application of the quantum entanglement concept can lead to reduced power consumption by reducing signalling overhead. This is based on the assumption that this technology is currently available, with a mathematical framework providing a model for behaviour.

6.1 Introduction

ORAN architecture is specifically designed to enhance the proximity of some fundamental network operations to users by means of the DU unit. This will result in a decrease in the duration of connection and a decrease in latency. Nevertheless, the implementation of more DU servers results in an increase in the overall PC of the network. Furthermore, the virtualization process used within the CU and DU introduces an additional latency, as each virtual computer contends with others to process the data of its respective users [62]. Therefore, it is essential to prioritise the reduction of PC in order to achieve optimal efficiency.

This research presents a three-phase solution within the context of ORAN. The proposed approach entails the use of load balancing techniques by integrating principles from quantum physics into classical communication systems, specifically by using the theory of entanglement. The latter facilitates the production of several units of quantum information by the utilisation of a single unit of classical information. This method aims to conserve power within the cloud infrastructure. Second, we have put out a streamlined yet efficient PC model for the ORAN architecture, which aims to accurately depict the variations in the volume of traffic handled by the server. Furthermore, it demonstrates the manner in which PC can be characterised inside ORAN system. Thirdly, by utilising this model, we are able to form an EE optimisation problem. It was established with the objective of selecting the least energy-efficient servers for the load balancing process established in the first phase; for the benefit of the server users. The presence of nonlinearity in the objective function necessitated the utilisation of the Lagrange multiplier method as a mathematical approach to solving the problem. In this study, the 'fmincon' method was employed to address the problem numerically. Specifically, two strategies were utilised and subsequently compared. Two algorithms have been considered for solving the optimisation problems: sequential quadratic programming (SQP) and the active-set algorithm. Both of these methods are specifically developed to address objective functions that exhibit characteristics of both nonlinearity and constraint behaviours.

6.2 Quantum based model

The suggested method involves leveraging entanglement to facilitate the exchange of load balancing signalling information amongst the servers using photons as shown in Fig 6.1. Traditionally, the act of signalling between servers has been associated with significant costs in terms of PC and time [163]. Therefore, the suggested quantum approach provides a progressive means of information sharing across servers. In the proposed method, every server is assigned a photon, which is subsequently transformed into classical information. The state of the photon is then modified in order to disseminate this information simultaneously to the servers of interest. Traditionally, this necessitates the transmission of signalling information from each server to every other server individually to track the processing availability, resulting in increased PC and network latency. In order to achieve load balancing across the CU servers, we have measured the EE of the server's users as our criteria to decide the load balancing servers. Hence, it is essential to consider two primary metrics in this process: data rate and PC.

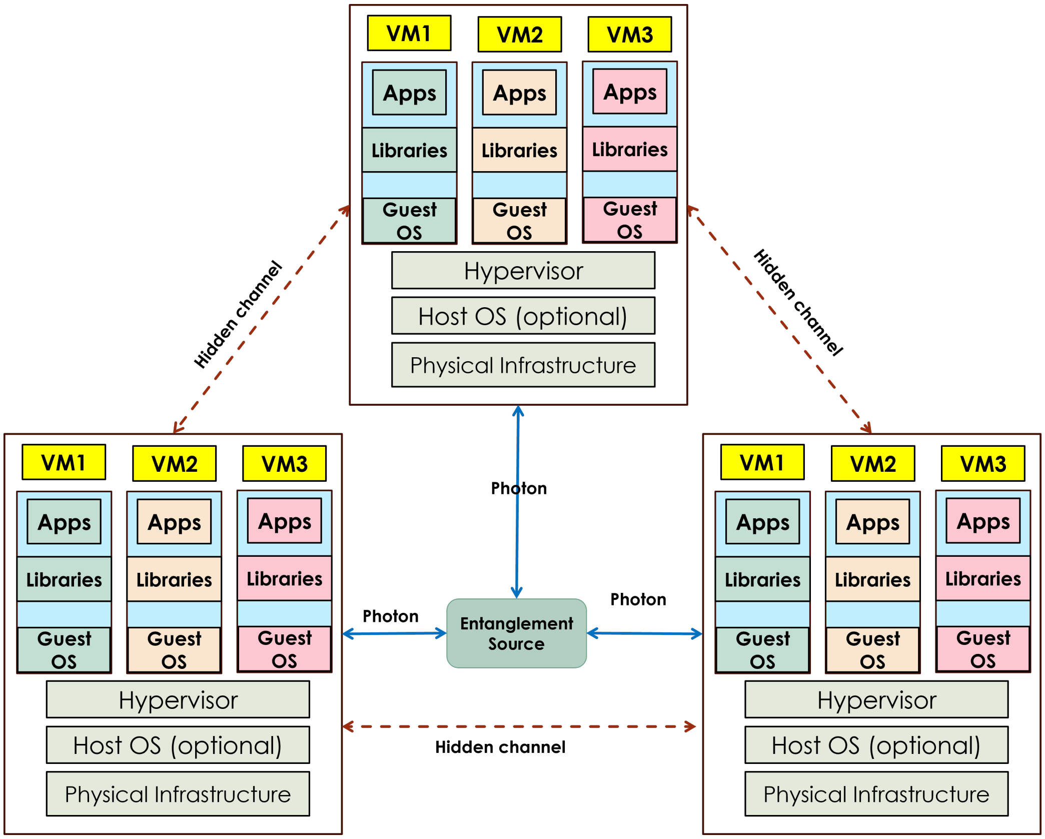

Fig. 6.1. Proposed quantum based load balancing in ORAN.

6.3 PC model

It was assumed the PC of the network consists of cloud infrastructure that contains CU servers, DUs, RUs. In ORAN, there are many VMs that are responsible for enlarging the PC of the server up from its initial value to its maximum. Not only that, each VM is responsible for processing many resource blocks (RB), translated to bits, at each time slot. Moreover, the initial PC value of the devices is different amongst the servers. This usually based on the network vendor, manufacturer and device's characteristic. On top of that, it is not perfectly known how much power the software itself consumes, driving the idle power of the hardware [193]. Hence, offering detailed power model is a prerequisite to include the aforementioned parameters.

The modulation type, denoted as x governs the number of bits L_x in each resource element, for example, in QPSK, the number of bits is $L_x = 2$. In 16-QAM, $L_x = 4$, following the constellation of the modulation type, where $mod_x = 2^L$, is the number of bits in the constellation diagram. This number is multiplied to number of sub-carriers in the RB N_{sc}^{rb}, and number of symbols in the RB N_{sym}^{rb} to measure the number of bits in each RB, as follows:

$$L_{rb}^x = L_x \times N_{sc}^{rb} \times N_{sym}^{rb} \tag{6.1}$$

To obtain the number of bits in all the RBs, the L_{rb}^x is multiplied by the number of RBs (N_rb), as follows:

$$L_{RB}^x = L_{rb}^x \times N_{rb} \tag{6.2}$$

On the other side, the power given in the RB is equivalent to

$$P_{rb} = \frac{P_{NB}}{N_{rb}} \tag{6.3}$$

where P_{NB} and N_{rb} denote the power transmitted of the nodeB and the number of RBs, respectively. Note this is the power consumed in one time slot T_s. If we assume the total time of transmission is T, where $T = T_s \times N_{rb}$ where the transmission is discontinuous. In case of continuous transmission, $T = \int_{t_o}^{t_l} dt$, where t_o and t_l are the beginning and ending time of the transmission. Hence, the time averaged power allocated to the RB P_{rb}^T over a period of time T, is given by:

$$P_{rb}^T = P_{rb} \times T \tag{6.4}$$

Which is assumed the total consumption of the VMs when processing a number of RBs. For one time slot, the VMs' PC is given by:

$$P_{vm} = \gamma P_{rb} \times N_{rb} \tag{6.5}$$

Where γ denotes the share of the VM software of the total PC of the server. For example, if the hardware consumes 70%, then the software is 30%, where P_{vm} impacts. For a group of VMs, the PC can be modelled as:

$$P_{VM} = P_{vm} \times N \tag{6.6}$$

where N is the total number of VMs.

6.3.1 CU PC

Modelling the PC of the CU can mainly based on the existed VMs. However, the change in the PC of the CU varies according to the change in the number of VMs (N) and change of the time t at which these VMs are existed. It is worth noting that at any time instance, there are some VMs that are installed and increased the PC P_c within the server, and some of the VMs that terminating processing the UE's data and spared some power P_r to the total CU server PC P_{CU}. meaning, the more spared power, the server can adapt more VMs and exploit the total consumption P_CU. Hence, the P_r is a function of P_CU, i.e., $P_r \propto P_{CU}$ or $P_r = \delta P_{CU}$, hence:

$$\frac{\partial P_{CU}}{\partial t} = P_c - P_r \tag{6.7}$$

or

$$\frac{\partial P_{CU}}{\partial t} = P_c - \delta P_{CU} \tag{6.8}$$

That is

$$\frac{\partial P_{CU}}{P_c - \delta P_{CU}} = \partial t \tag{6.9}$$

Integrating both sides reveals:

$$\frac{\ln(P_c - \delta P_{CU})}{-\delta} = t + c \tag{6.10}$$

where c is the constant of integration. The above expression can be written as :

$$P_c - \delta P_{CU} = e^{-\delta t} e^{-\delta c} = Ce^{-\delta t} \tag{6.11}$$

If the initial condition is applied, the P_{CU} and t are replaced by P^o_{CU} and t_o, respectively. Hence:

$$C = (P_c - \delta P^o_{CU})e^{\delta t_o} \tag{6.12}$$

Substitute Eq. 6.12 in 6.11, yields:

$$P_{CU}(t) = \frac{P_c}{\delta} + (P^o_{CU} - \frac{P_c}{\delta})e^{-\delta(t-t_o)} \tag{6.13}$$

it is worth mentioning that the initial consumption is represented by P^o_{CU}. In addition, the model produces P_c/δ, which is the maximum value of consumption. To model the effect of VMs up on the CU consumption, replacing the time with number of VMs produces the same effect upon the P^o_{CU}, as follows:

$$P_{CU}(N) = \frac{P_c}{\delta} + (P^o_{CU} - \frac{P_c}{\delta})e^{-\delta(N-N_o)} \tag{6.14}$$

where N_o is the initial number of VMs.

6.3.2 DU PC

The DU server serves as a crucial intermediary component connecting the CU server and the RU unit. The responsible components encompass several essential activities, namely the physical layer, media access control layer, and transport layer. The rationale behind the proximity of these functions to the users is to minimise call activation time by avoiding communication with distant cloud centres. The assumption was made that the DU server is also virtualized. Each VM is tasked with the responsibility of establishing communication with a single radio unit. In order to determine the PC of the DU server, we employed a similar methodology to that utilised for the CU server, as outlined below:

$$P_{DU}(t) = \frac{P_c}{\delta} + (P^o_{DU} - \frac{P_c}{\delta})e^{-\delta(t-t_o)} \tag{6.15}$$

and

$$P_{DU}(N) = \frac{P_c}{\delta} + (P^o_{DU} - \frac{P_c}{\delta})e^{-\delta(N-N_o)} \tag{6.16}$$

6.3.3 Radio Unit

The radio unit is connected to the UEs from one side, and to the DU from the other side. It was assumed the RU is a bare device and not virtualized. It includes two main units, the radio frequency and the power amplifier unit.

The PC of the power amplifier can be evaluated by considering the maximum transmission power of the NodeB.

$$P_{PA} = \frac{P_{NB}}{\eta} \tag{6.17}$$

where η denotes the PA's efficiency. It is noted that this unit is affected by the number of transmitted RBs, if we substitute $P_{NB} = P_{rb} \times N_{rb}$, Equation 6.17, is written as:

$$P_{PA} = \frac{P_{rb} \times N_{rb}}{\eta} \tag{6.18}$$

This means the more bandwidth allocated to the system, the more PC of the power amplifier. In addition, the radio frequency unit P_{RF} is slightly affected by the transmitted RBs [172].

Hence, the radio unit PC P_{RU} is the summation of the radio frequency and power amplifier units.

$$P_{RU} = P_{RF} + P_{PA} \tag{6.19}$$

6.4 Quantum PC

Modelling the entanglement-wise principal component is of utmost importance. In practical terms, a laser beam is employed to excite a nonlinear crystal known as a beta barium borate (BBO) crystal. The semiconductor laser induces the generation of photon pairs in the crystal, wherein the frequency of each twin photon is halved and the wavelength is doubled compared to the initial laser beam. The PC associated with the laser is represented as P_{laser}, while the PC connected to the detector is designated as P_{det}. Similarly, the PC linked to the driver is indicated as P_{driver}. The expression for the quantum entanglement's PC, denoted as P_{ent}, is provided as follows:

$$P_{ent} = P_{laser} + P_{det} + P_{driver} \tag{6.20}$$

If we assume the traditional PC is:

$$P_{traditional} = \frac{(P_{CU} + P_{DU})}{P_{loss}} + \frac{P_{RU}}{P_{AcDc}} \tag{6.21}$$

where P_{loss} denotes the PC of the AC-AC P_{AC}, AC-DC P_{DC} and cooling P_{cool}, where

$$P_{loss} = P_{AC} + P_{DC} + P_{cool} \tag{6.22}$$

The cooling PC is excluded in the RU unit, as.

$$P_{AcDc} = P_{AC} + P_{DC} \tag{6.23}$$

The quantum PC $P_{quantum}$ is given by adding P_{ent} to the quantum case while deducting the power saving value P_{save} from the traditional consumption, as follows:

$$P_{quantum} = (P_{traditional} - P_{save}) + P_{ent} \tag{6.24}$$

To formulate P_{save}, the PC of the messages from the server of interest to other servers is calculated. Since the number of bits in each signalling message is known parameter, the number of RBs in each message to be sent to server s; using modulation type x; can be evaluated as

$$N_{rb,s,x}^{msg} = \frac{L_{msg,s,x}}{L_{rb}^{s,x}} \tag{6.25}$$

where $L_{msg,s,x}$ denotes the number of bits contained in the message that is sent to server s using x type of modulation. To obtain the power contained in the message, $N_{rb,s,x}^{msg}$ is multiplied by the PC of RB, as follows:

$$P_{rb,s,x}^{msg} = N_{rb,s,x}^{msg} \times P_{rb} \tag{6.26}$$

Hence, P_{save} is the power consumed for round-trip transmitting the messages with $P_{rb,s,x}^{msg}$ to a total number of servers S, i.e.:

$$P_{save} = S \times P_{rb,s,x}^{msg} \tag{6.27}$$

6.5 Selecting Specific servers

According to [194], the process of generating entanglement is facilitated when a smaller number of photons are created. This phenomenon has the potential to impact the necessary optical and electrical equipment, as well as the acquisition of exceptionally pure photons, a high-fidelity system, and more prominent polarisation states. The current task necessitates the selection of servers to which these photons will be transmitted, in order to facilitate load balancing in a manner that minimises the amount of created photons. This contributes to the reduction in the quantity of photons created and mitigates the intricacy of the proposed system. The cloud infrastructure consists of a substantial quantity of servers, which can be partitioned into smaller subsets. Each subset can then be subjected to an optimisation procedure.

Suppose a cloud with servers set S, each server s is serving several users, where the total number of users is U, and $u_{s,n}$ is the users that are served by n-th VM in the server s. It is required to optimise the EE of these servers so as the servers with minimum EE are selected for the quantum load balancing. In general, the EE is given by $E = \dfrac{C}{PC}$ The power consumption, denoted as PC, can be represented by either $P_{traditional}$ or $P_{quantum}$ depending on the system being assessed. The selection method involves the utilisation of the traditional PC. Upon resolving the optimisation problem, the PC $P_{quantum}$ can be employed to facilitate the comparison between the pre-optimization and post-optimization states. The proposed optimization method has been implemented for single and two servers, following some assumptions regarding both cases, as follows:

1. The number of VMs that are allocated to each server is proportional to the number of users U. We assumed that $N_s = U_s/r$, where r is the share of the VMs is the server s. For example, if server 1 is connected to 4 UEs and server 2 is connected to 2 UEs and r=2, it means server 1 and 2 contain 2VMs and 1 VM, respectively. In addition, U_s denotes the total number of users within servers s; $U_s \in U$ and U is the total number of users.
2. The maximum consumption of a server is denoted as PCU.
3. The minimum achieved sum data rate is guaranteed via the C_{max} constraint.
4. The PC model is reformulated as a linear model to force the effect of the N_s and RBs variables with in both the PC and the data rate.
5. In case of two servers, the linear PC model is also converted to a linear model as this has the same effect on both servers.

In a single server, the EE is formulated as follows:

$$\begin{aligned} \max \quad & \frac{\sum_{u=1}^{U_s} B\ log(1 + \dfrac{Pt_{s,n,u} H_{s,n,u}}{N_o + I})}{P_{CU,s} + (N_s \times P_{vm}^{s,v,n})} \\ \text{s.t.} \quad & N_s \leq N \\ & P_{CU,s} + (N_s \times P_{vm}^{s,v,n}) \leq PCU \\ & C \geq C_{thr} \\ & N_s, P_{CU,s}, C \geq 0 \end{aligned} \tag{6.28}$$

If we assume $C = x$, $N_s = y$, $P_C U = z$, and substituting $N_s = U_s/r$, $P_{vm}^{s,v,n} = a$ the optimisation problem becomes:

$$\max \quad F(x,y,z) = \frac{rxy}{z + (ay)}$$

$$\begin{aligned} \text{s.t.} \quad & y \leq N \\ & z + ay \leq PCU \\ & x \geq C_{thr} \\ & x, y, z \geq 0 \end{aligned} \tag{6.29}$$

The type of this problem is nonlinear and hence, Lagrange multiplier solution is used to obtain the optimal values that maximise the problem. In which, the general expression is given by $\mathcal{L} = F(x,y) - \lambda\big(g(x,y) - c\big)$, where $g(x,y)$ is the constraint functions. The full solution can be found in Appendix A.

$$\mathcal{L} = \frac{rxy}{z + (ay)} + \lambda_1(N - y) + \lambda_2(C - C_{thr}) + \\ \lambda_3(PCU - z - y)$$

$$\begin{aligned} \partial L_x &= \frac{ry}{z + (ay)} + \lambda_2 \\ \partial L_y &= \frac{rx(z + ay) - (raxy)}{(z + (ay))^2} + \lambda_1 - \lambda_3 \\ \partial L_z &= \frac{-rxy}{(z + (ay))^2} + \lambda_3 \end{aligned}$$

Solving these equations results the following:

$$\lambda_1 = \frac{rx}{z+ay}$$
$$\lambda_2 = \frac{-rxy}{(z+ay)^2}$$
$$\lambda_3 = \frac{-ry}{z+ay}$$

Or can be solved with respect to x,y,z, as follows:

$$x = \frac{-z\lambda_3}{r+a\lambda_3} \leq C_{thr}$$
$$y = \frac{-z\lambda_3}{r+a\lambda_3} \leq N$$
$$z = \frac{y(\lambda_1 a - rx)}{rx + -\lambda_1} \leq PCU - ay$$

Given the values of a, r, N, PCU, C_{thr}, the values of x,y,z can be obtained.

In the first formulation, the EE problem is presented to maximize the server's itself, while in the second formulation, the EE of two servers is presented:

$$\text{Maximize:} \quad \frac{r \cdot y_1 \cdot x_1}{z_1 + a \cdot y_1} + \frac{r \cdot y_2 \cdot x_2}{z_2 + a \cdot y_2}$$

$$x_1 + x_2 \geq C_{\text{thr}}$$
$$y_1 + y_2 \leq N$$
$$z_1 + a \cdot y_1 \leq \text{PCU}_1$$
$$z_2 + a \cdot y_2 \leq \text{PCU}_2$$
$$z_1 + z_2 = \text{PCU}_{1,\text{init}} + \text{PCU}_{2,\text{init}}$$

$$\begin{aligned}\mathcal{L}(x_1, x_2, y_1, y_2, z_1, z_2, \lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5) = \frac{r_1 \cdot y_1 \cdot x_1}{z_1 + a_1 \cdot y_1} + \frac{r_2 \cdot y_2 \cdot x_2}{z_2 + a_2 \cdot y_2} + \\ \lambda_1 \cdot (x_1 + x_2 - C_{\text{thr}}) + \lambda_2 \cdot (N - y_1 - y_2) + \\ \lambda_3 \cdot (\text{PCU1} - z_1 - a_1 \cdot y_1) + \lambda_4 \cdot (\text{PCU2} - z_2 - a_2 \cdot y_2) + \\ \lambda_5 \cdot (z_1 + z_2 - (\text{PCU1ini} + \text{PCU2init}))\end{aligned}$$

where

$$\partial L_{x1} = \frac{r_1 \cdot y_1}{z_1 + a_1 \cdot y_1} + \lambda_1 = 0$$

$$\partial L_{x2} = \frac{r_2 \cdot y_2}{z_2 + a_2 \cdot y_2} + \lambda_1 = 0$$

$$\partial L_{y1} = \frac{r_1 \cdot x_1 \cdot (z_1 + a_1 \cdot y_1) - r_1 \cdot y_1 \cdot a_1 \cdot x_1}{(z_1 + a_1 \cdot y_1)^2} - \lambda_2 - \lambda_3 \cdot a_1 = 0$$

$$\partial L_{y2} = \frac{r_2 \cdot x_2 \cdot (z_2 + a_2 \cdot y_2) - r_2 \cdot y_2 \cdot a_2 \cdot x_2}{(z_2 + a_2 \cdot y_2)^2} - \lambda_2 - \lambda_4 \cdot a_2 = 0 \tag{6.30}$$

$$\partial L_{z1} = -\frac{r_1 \cdot y_1 \cdot x_1}{(z_1 + a_1 \cdot y_1)^2} - \lambda_3 + \lambda_5 = 0$$

$$\partial L_{z2} = -\frac{r_2 \cdot y_2 \cdot x_2}{(z_2 + a_2 \cdot y_2)^2} - \lambda_4 + \lambda_5 = 0$$

Subsequently, the solution of these equations can be found in Appendix B, where $\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5, x_1$, the optimisation variables are obtained. In addition, the proof of convexity/concavity can be found in Appendix C.

6.6 Results and Discussion

Fig. 6.2 shows the PC comparison of the CU server with respect to different values of the P_c, while considering P_{CU}^o is 60W, $t_o = 0$, and t is up to 2 seconds. The model shows that the more consumed power P_c, the steady state of the model becomes ideal due to the overcoming behaviour of P_c value over the static value. In Addition, δ has been taken as 0.4. However, these values can be adjusted based on the different characteristic of the servers. Meaning, given the device's specifications, the maximum and static consummations' can be allocated in the model to present the mediate PC values with respect to the time or number of virtual machines.

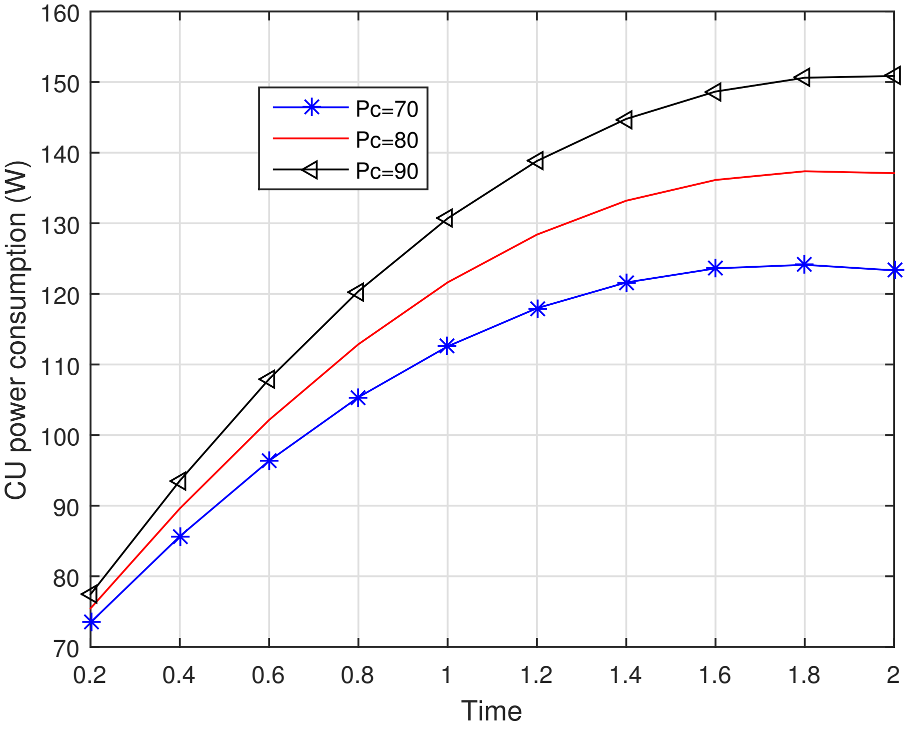

Fig. 6.2. Effect of P_c on the PC of the CU server.

In the next Figure, the P_c is fixed to 90W, while the values of the P^o_{CU} are changed. It is shown that the value of the static consumption only affects the total value of the consumption. The other assumption are kept the same as Fig 6.2.

The following Fig. 6.4 shows the effect of VMs on the PC of the server. Note that such server might be considered as CU or DU server. The values of P_c and P^o_{CU} are fixed to 80W and 70W, respectively. This figure shows the different behaviours of the model with respect to different values of α, which reflects on different servers specifications.

To linearize and simplify the model, we have assumed the model contains the static and dynamic consumption, such as $P^o_{CU,s} + (N_s \times P^{s,v,n}_{vm})$. Not to forget it is

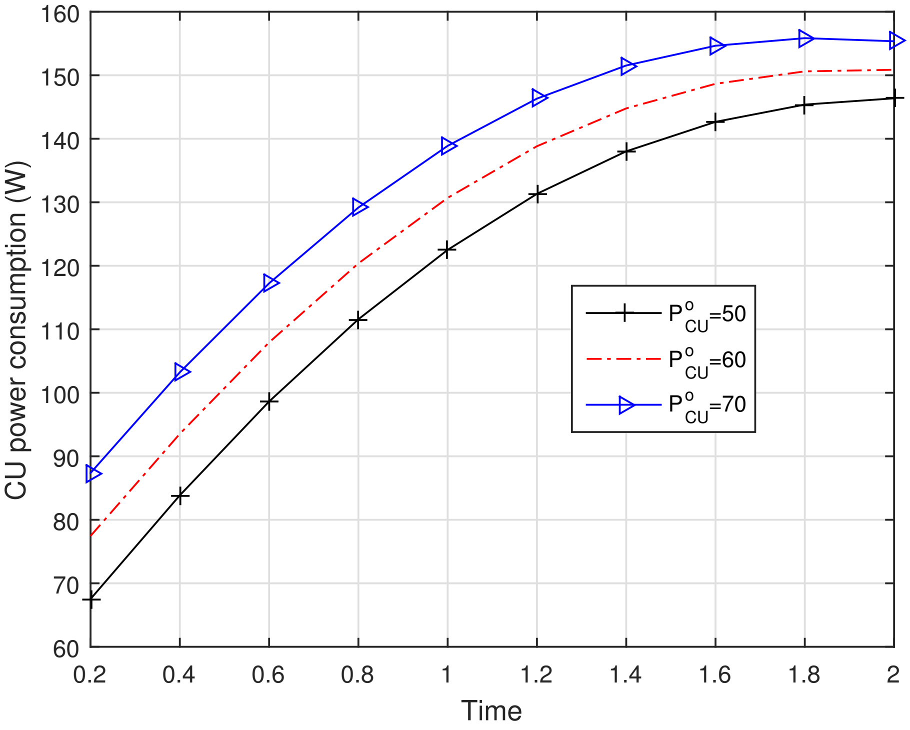

Fig. 6.3. Effect of P^o_{CU} on the PC of the CU server.

Table 6.1

Model Parameters III

Factor	Value	Unit	Factor	Value	Unit
P_{bit}	0.3	W	L_{bit}	20	bits
N_{rb}	2	RBs	C_{thr}	100	bps
P^o_{CU1}	70	W	P^o_{CU2}	70	W
$PCU1$	100	W	$PCU2$	110	W
r_1	0.6	-	r_2	0.4	W
P_{driver}	1	W	P_{synch}	1	W
P_{delta}	0.4	-	P	1	W

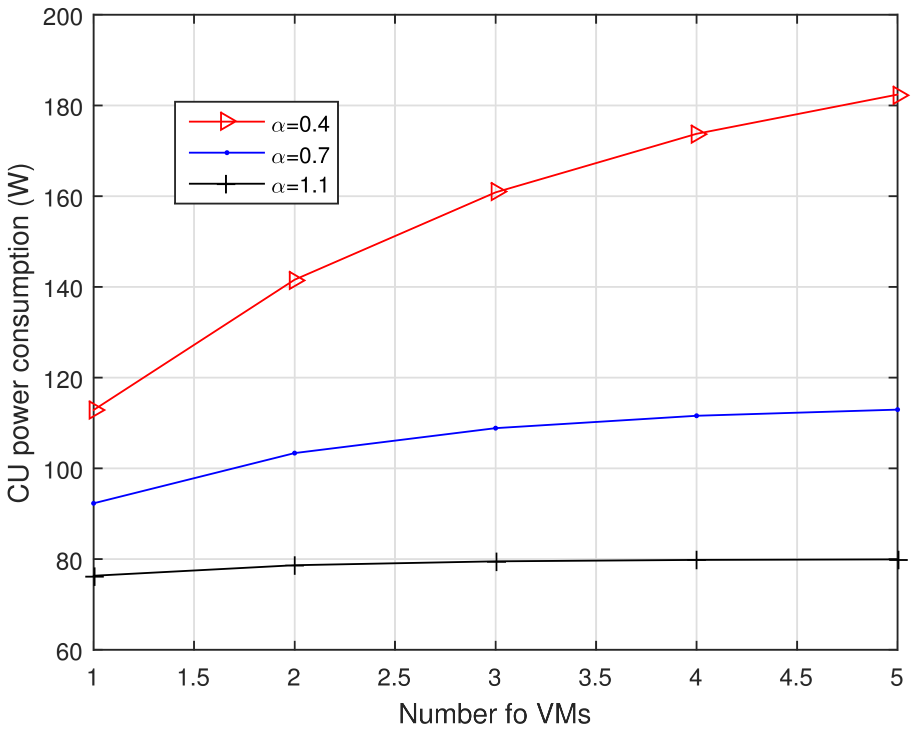

Fig. 6.4. Effect of number of virtual machines on the PC of the CU server.

crucial to decide the number of VMs to be placed in one server based on the coming load, which is also an optimization problem that contributes to lessen the power consumed by each server. However, in this work, to hold the objective function being dependant on the number of VMs in both the channel capacity and the PC. We have assumed that the number of VMs in the server is based on its connected users and a weighting parameter r, that is decidable and agreed amongst the network operators based on their ORAN policy agreement. Subsequently, when the value of r is decided, it affects the value of VMs, which affects the value of E. Furthermore, the optimisation problem is further constrained by the involvement of many vendors and operators. This is because the sharing policy might either involve complete sharing of cloud resources or be based on a weighing technique. In either case, it is necessary to develop new optimisation algorithms that prioritise the following issues:

1. The cloud is shared by network suppliers or operators. The remedy to the problem can vary depending on the individual's specific settings of PC for their devices.

2. The network operator allocates a portion of the cloud resources, either small or large, depending on its projected user base or through a pay-as-you-go approach. This implies that the values of PCs undergo temporal fluctuations.

3. In the ORAN architecture, many operators cohabit within a shared server environment. The current inquiry pertains to whether all of these operators are in consensus over the utilisation of virtualization technology to facilitate connectivity for their subscribers. In this scenario, it is anticipated that there will be an increase in latency. Therefore, the inclusion of a delay restriction can be incorporated into the optimisation problem.

The 'fmincon' solver has been employed as the preferred optimisation method due to its ability to effectively handle nonlinear objective functions and constraints, encompassing both inequality and equality constraints. The active-set approach and sequential quadratic programming (SQP) are two algorithms that are utilised and compared. The SQP method is a widely used approach for addressing nonlinear constrained optimisation issues. The method employed is iterative in nature, wherein the subproblem is resolved within the confines of the feasible zone. Simultaneously, the overarching problem is progressively approximated through the utilisation of a quadratic model. This method is well recognised as a highly efficient and effective approach for addressing complex nonlinear optimisation problems that involve both equality and inequality requirements.

The SQP algorithm integrates components from both Newton's approach, which involves second-order optimisation techniques, and the Lagrange multipliers method, which addresses the handling of constraints. The algorithm commences with initialising and selecting an initial feasible point $x(0)$, followed by setting the iteration counter k to zero. Next, we will proceed with the construction of a quadratic approximation for both the objective function and the constraint, centred around the point denoted as $x(k)$. Next, the quadratic sub-problem is solved in order to choose a search direction $p(k)$ that minimises the objective function. Subsequently, the point can be updated by employing the formula $x(k+1) = x(k) + \gamma p(k)$, where γ represents a step size that ensures both feasibility and progress towards the optimal value. Subsequently, the convergence can be assessed by employing specified criteria, such as evaluating constraint violation or monitoring changes in the objective function. Lastly, the counter k is incremented by adding 1, denoted as $k = k + 1$, and the process returns to step 2.

In contrast, the active set approach employs the identical criteria as the SQP method, however with convergence achieved when the solution to the sub-problem satisfies the Karush-Kuhn-Tucker (KKT) conditions. The two methods have been compared in terms of their performance using the given parameters. Figure 6.5 shows the different decision variables with respect to the objective function values of the SQP.

In Figure 6.6, the decision variables were depicted in relation to the objective function when employing the Active-set algorithm.

In general, the SQP algorithm out perform the Active-set method. A comparison has been made between the two methods while adjusting some of the decision variables, such as C_{thr}, as shown in Fig. 6.7, when considering 20 users.

The convergence characteristics of the optimisation algorithm can be impacted by a range of aspects, encompassing the starting solution, the selection of the optimisation strategy, the attributes of the problem under consideration, and the particular configurations of the optimisation approach. Certain algorithms may demonstrate rapid convergence, while others may require a larger number of iterations to achieve a satisfactory outcome. In the real implementation, the suggested comparison may not produce favourable outcomes within the specific context of the requested work. Irrespective of the specific method utilised, it is feasible to ascertain the server exhibiting the lowest or maximum energy efficiency (EE), albeit with differing numbers for each approach. While heuristics can provide alternative approaches for compar-

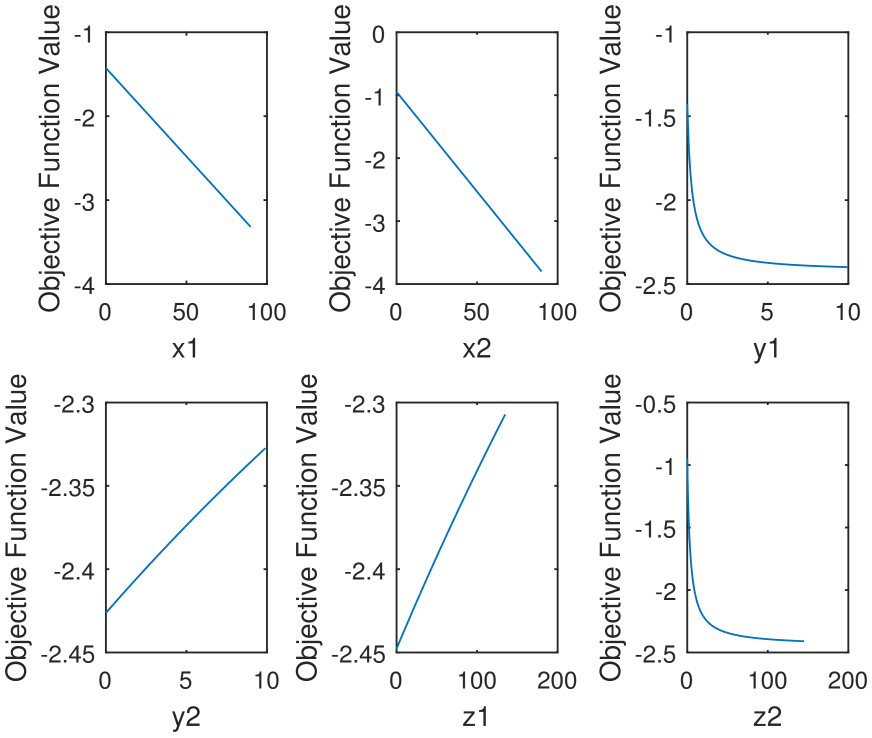

Fig. 6.5. Decision parameters with respect to the objective value using SQP algorithm.

ing algorithms, the selection criteria for servers does not necessitate comprehensive knowledge of which method yields higher energy efficiency. Rather, it requires selecting certain servers. Therefore, conducting more comparisons is only time-consuming.

6.7 Conclusions and future trends

The installation of a substantial quantity of virtual machines, wherein each virtual machine consumes processing resources, leads to an expansion regarding the PC. Therefore, it is imperative to optimise the allocation of VMs across several operators in order to provide a highly efficient network. This study presents a power model that examines the impact of various network factors, notably the time factor and the

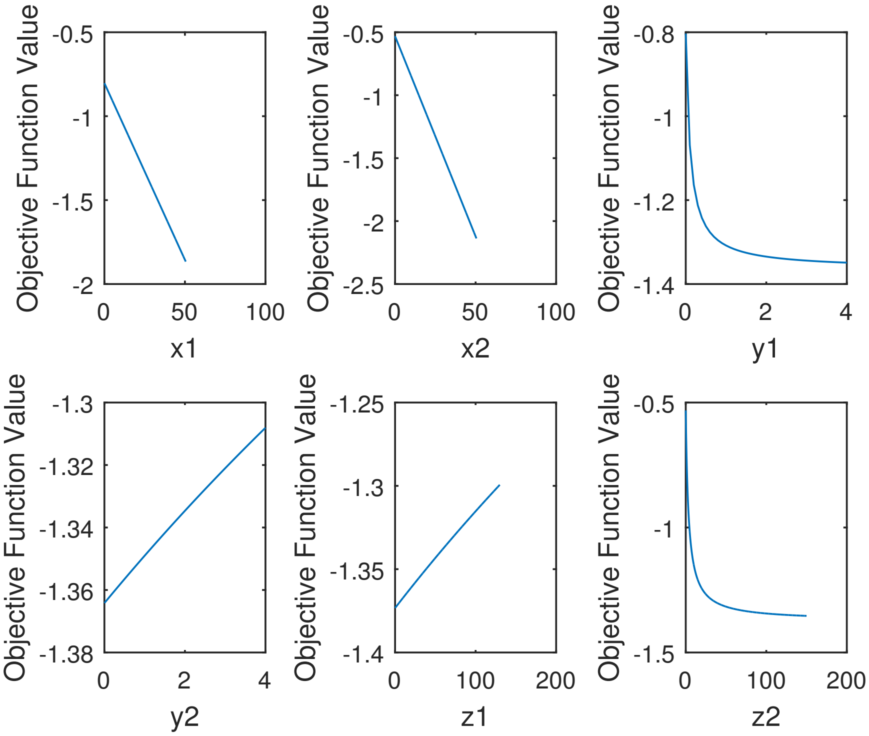

Fig. 6.6. Decision parameters with respect to the objective value using Active-set algorithm.

quantity of VMs, on power consumption. This model is considered new because to its focus on the ORAN architecture. Furthermore, it is characterised by its simplified and realistic approach, distinguishing it from other models discussed in existing literature, which may involve complex functions, hardware, or software components. This model employed in the EE maximisation problem incorporated the proposed quantum entanglement approach. The process involves a comparison of the servers in order to determine the most suitable load balancing servers from the available options within the ORAN cloud. The optimisation problem was solved with the Lagrange multiplier technique and afterwards numerically solved employing the SQP approach, which demonstrated superior performance compared to the Active-set methods.

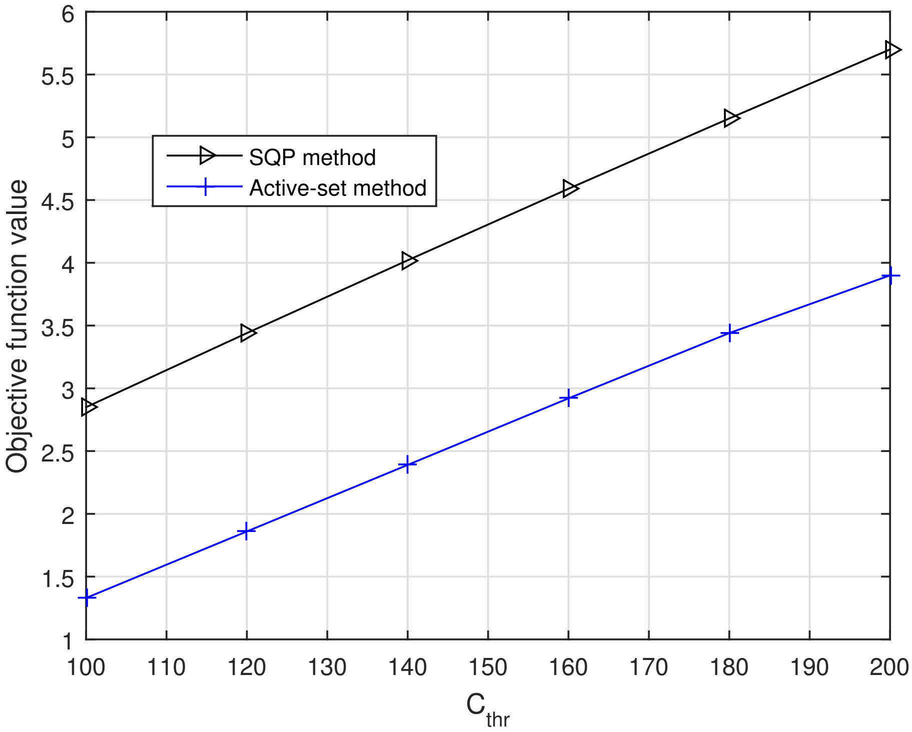

Fig. 6.7. The threshold data rate with respect to the objective value using SQP and Active-set algorithms.

In the future, the coexistence of several network operators within ORAN servers necessitates the optimisation of resource allocation to determine which operator can derive the greatest advantage from the servers' resources at a certain time window. The determination of various weights, such as bandwidth and power, is contingent upon the cost factor, which is ultimately influenced by the concept of infrastructure as a service.

Chapter 7: Conclusions and Future Works

7.1 Conclusion

1. This thesis assesses the implications associated with scaling up the number of VMs and RBs deployed on ORAN servers. Virtualization has been found to reduce network PC when compared to the legacy CRAN. However, ORAN achieves higher PC scores as a result of utilising DU servers. Nevertheless, the research fails to take into account the potential advantages for vendors in terms of PC and latency. However, this study presented a generalized PC model that can quantify the PC gain in ORAN and compare it with the legacy architectures.

2. The incorporation of quantum entanglement in cellular communications yields enhanced performance of the X2-AP protocol, resulting in a notable decrease of 40% in both delay and energy consumption. This reduction is observed as the quantity of entangled photons grows. These percentages can also be attained while considering the signaling mechanism between the CUs and DUs. However, the aforementioned network factors, including the number of entangled photons, number of transmitted messages, system bandwidth, and number of CUs and DUs, have an impact on these percentages.

3. The deployment of a significant number of virtual machines results in an increase in the capacity of the personal computer. Hence, it is crucial to optimise the distribution of VMs across many operators to ensure the provision of a network that operates with high efficiency. The EE maximisation problem has

been formulated using the PC model. This approach entails conducting a comparative analysis of the EE of the servers to identify the most appropriate load balancing servers among the servers present inside the ORAN. The optimisation problem was resolved with the Lagrange multiplier method and afterwards solved numerically. This approach has the potential to mitigate the expenses related to the transmission of signals among various organisations inside the cloud infrastructure.

7.2 Future works

1. Virtualisation is a technique employed to achieve logical segregation of distinct network services or operations. Each individual slice has the capability to implement its own resource allocation policies, regardless of the vendor being used. This enables the allocation of resources in accordance with the unique requirements of each segment, without being restricted to a singular vendor's solution.

2. The implementation of a policy-driven resource allocation system is vital, this entails the establishment of regulations that delineate the appropriate allocation of resources in accordance with diverse network conditions, quality of service prerequisites, and service-level agreements. The aforementioned regulations possess the characteristic of being vendor-agnostic, allowing for their consistent application throughout the network.

3. Quantum entanglement can be employed not only among the RRHs, but also between the RRHs and the cloud centre. This process entails the efficient update of the cloud, MME, and SGW without incurring any time-related expenses. It was anticipated that the implementation of this strategy will yield enhancements in service quality with respect to both latency and power consumption. Nevertheless, it is imperative to do a thorough analysis of the simultaneous trade-offs in terms of both cost and system complexity. The aforementioned

objective can be achieved by the utilisation of artificial intelligence and quantum computing techniques, which facilitate the management of several aspects of the photon transmission process. These include controlling the transmission and reception of photons, purifying the polarisation states of the photons, updating the participants involved in the handover, and rectifying errors in undiscovered photons. Moreover, enhancing the efficiency of the suggested approach to encompass RRHs that are not linked to the same cloud centre could introduce additional complexity.

4. The distinctive characteristic of entangled particles possesses the capacity to significantly transform network synchronisation. The technology has the capability to facilitate highly accurate clock synchronisation between two distant locations within the ORAN network. Moreover, entangled particles provide inherent security as a result of quantum indeterminacy, hence guaranteeing confidentiality and resistant timing information against tampering.

APPENDICES

Appendix A: Solution for Lagrange multiplier

Solving the first objective function with respect to x, y, z.

Given equations:

1. $(ry)/(z + ay) + \lambda_3 = 0$

2. $(rx(z + ay) - rxya)/((z + ay)^2) - \lambda_1 - \lambda_2 a = 0$

3. $-(rxy)/((z + ay)^2) - \lambda_2 = 0$

Equation 1:

$$(ry)/(z + ay) + \lambda_3 = 0$$

Multiplying both sides by $(z + ay)$:

$$ry + \lambda_3(z + ay) = 0$$

Expanding:

$$ry + 3\lambda_3 z + \lambda_3 ay = 0$$

Isolating y:

$y(r + \lambda_3 a) = -\lambda_3 z$

Dividing by $(r + \lambda_2 a)$:

$y = -(\lambda_3 z)/(r + \lambda_3 a)$

Equation 2:

$(rx(z + ay) - rxya)/((z + ay)^2) - \lambda_1 - \lambda_2 a = 0$

Multiplying both sides by $((z + ay)^2)$:

$rx(z + ay) - rxya - \lambda_1((z + ay)^2) - \lambda_2 a((z + ay)^2) = 0$

Expanding:

$rxz + rxay - rxya - \lambda_1(z^2 + 2ayz + a^2y^2) - \lambda_2 a(z^2 + 2ayz + a^2y^2) = 0$

Collecting for x:

$rxz - rxya = \lambda_1(z^2 + 2ayz + a^2y^2) + \lambda_2 a(z^2 + 2ayz + a^2y^2) - rxay$

Isolating x:

$rx(z - ay) = (\lambda_1 + \lambda_2 a) * (z^2 + 2ayz + a^2y^2)$

Dividing by $(z - ay)$:

$x = ((\lambda_1 + \lambda_2 a)(z^2 + 2ayz + a^2y^2))/(r(z - ay))$

Equation 3:

$-(rxy)/((z + ay)^2) - \lambda_2 = 0$

Multiplying both by $((z + ay)^2)$:

$-rxy - \lambda_2((z + ay)^2) = 0$

Expanding:

$-rxy - \lambda_2(z^2 + 2ayz + a^2y^2) = 0$

Isolating y:

$y(-rx - \lambda_2 a^2) = -\lambda_2(z^2 + 2ayz)$

Dividing by $(-rx - \lambda_2 a^2)$:

$y = -(\lambda_2(z^2 + 2ayz))/(-rx - \lambda_2 a^2)$

Hence, the solutions for x and y are as follows:

$x = ((\lambda_1 + \lambda_2 a)(z^2 + 2ayz + a^2y^2))/(r(z - ay))$

$y = -(\lambda_3 z)/(r + \lambda_3 a)$

Solving for z

1. First equation:

$\lambda_3 = -(ry)/(z + ay)$

2. Second equation:

$(rx(z + ay) - rxya)/((z + ay)^2) = \lambda_1 + \lambda_2 a$

3. Third equation:

$\lambda_2 = -(rxy)/((z + ay)^2)$

Eliminating λ_2 from Eqs. (2) and (3):

$(rx(z + ay) - rxya)/((z + ay)^2) = \lambda_1 - (rxy)/((z + ay)^2)$

Cross-multiplying to ignore the denominators:

$(rx(z + ay) - rxya) = \lambda_1(z + ay) - (rxy)$

Distributing λ_1:

$rxz + rxay - rxya = \lambda_1 z + \lambda_1 ay - rxy$

Rearranging and isolating z:

$rxz - \lambda_1 z = -rxay + \lambda_1 ay$

Factoring out z:

$$z(rx - \lambda_1) = y(\lambda_1 a - rx)$$

$$z = (y(\lambda_1 a - rx))/(rx - \lambda_1)$$

Appendix B: Solution of the optimization problem

Solving for the second objective function, solving for $x_1, x_2, y_1, y_2, z_1,$ and z_2, the variables are isolated in each equation and then solving the resulting equations simultaneously, as follows:

1. First equation:

$\frac{r_1 y_1}{z_1+a_1 y_1} + \lambda_1 = 0$

Solving for y_1:

$y_1 = -\frac{\lambda_1(z_1+a_1 y_1)}{r_1}$

Rearranging and isolating y_1:

$(1 + \frac{\lambda_1 a_1}{r_1})y_1 = -\frac{\lambda_1 z_1}{r_1}$

$y_1 = -\frac{\lambda_1 z_1}{r_1(1+\frac{\lambda_1 a_1}{r_1})}$

2. Second equation:

$\frac{r_2 y_2}{z_2+a_2 y_2} + \lambda_1 = 0$

Solve for y_2:

$y_2 = -\frac{\lambda_1(z_2+a_2 y_2)}{r_2}$

Rearranging and isolating y_2:

$(1 + \frac{\lambda_1 \cdot a_2}{r_2})y_2 = -\frac{\lambda_1 z_2}{r_2}$

$y_2 = -\frac{\lambda_1 \cdot z_2}{r_2(1+\frac{\lambda_1 \cdot a_2}{r_2})}$

3. Third equation:

$\frac{r_1 x_1(z_1+a_1 y_1)-r_1 y_1 a_1 x_1}{(z_1+a_1 y_1)^2} - \lambda_2 - \lambda_3 a_1 = 0$

Solving for x_1:

$x_1 = \frac{r_1 y_1 a_1 x_1 - r_1 x_1 z_1 + \lambda_2(z_1+a_1 y_1)^2}{r_1 y_1 a_1}$

Rearranging and isolating x_1:

$(1 - \frac{\lambda_2}{r_1 y_1 a_1})x_1 = \frac{\lambda_2(z_1+a_1 y_1)^2 - r_1 x_1 z_1}{r_1 y_1 a_1}$

$x_1 = \frac{\lambda_2(z_1+a_1 y_1)^2 - r_1 x_1 z_1}{r_1 y_1 a_1(1-\frac{\lambda_2}{r_1 y_1 a_1})}$

4. Fourth equation:

$\frac{r_2x_2(z_2+a_2y_2)-r_2y_2a_2x_2}{(z_2+a_2y_2)^2} - \lambda_2 - \lambda_4a_2 = 0$

Solving for x_2:

$x_2 = \frac{r_2y_2a_2x_2-r_2x_2z_2+\lambda_2(z_2+a_2y_2)^2}{r_2y_2a_2}$

Rearranging and isolating x_2:

$(1 - \frac{\lambda_2}{r_2y_2a_2})x_2 = \frac{\lambda_2(z_2+a_2y_2)^2-r_2x_2z_2}{r_2y_2a_2}$

$x_2 = \frac{\lambda_2(z_2+a_2y_2)^2-r_2x_2z_2}{r_2y_2a_2(1-\frac{\lambda_2}{r_2y_2a_2})}$

5. Fifth equation:

$-\frac{r_1y_1x_1}{(z_1+a_1y_1)^2} - \lambda_3 + \lambda_5 = 0$

Solve for z_1:

$z_1 = \frac{r_1x_1y_1}{(1-\frac{\lambda_3}{r_1y_1})}$

6. Sixth equation:

$-\frac{r_2y_2x_2}{(z_2+a_2y_2)^2} - \lambda_4 + \lambda_5 = 0$

Solving for z_2:

$z_2 = \frac{r_2x_2y_2}{(1-\frac{\lambda_4}{r_2y_2})}$

Appendix C: Proof of convexity/concavity

In order to examine the convexity of the propsed problem, it is necessary to evaluate the second derivatives of the objective function with respect to the variables x_1, x_2, y_1, y_2, z_1, and z_2.

1-The second derivative of the function f with respect to x_1 is equal to zero.

2-The second derivative of the function f with respect to x_2 is equal to zero. Given that the second derivative remains constant and non-negative, this does not yield definitive conclusions on the convexity of the function.

3-The second derivative of the function f with respect to y_1 can be expressed as follows:

$$\frac{\partial^2 f}{\partial y_1^2} = -2r_1x_1(z_1 + a_1y_1)$$

The negative value of the second derivative implies a concave.

4-The second derivative of the function f with respect to y_2 can be expressed as follows:

$$\frac{\partial^2 f}{\partial y_2^2} = -2r_2x_2(z_2 + a_2y_2)$$

The negative value of the second derivative suggests a concave shape.

5-The second derivative of the function f with respect to z_1 can be expressed as follows:

$$\frac{\partial^2 f}{\partial z_1^2} = 2r_1y_1x_1(2a_1y_1 - z_1)$$

6-The second derivative of the function f with respect to z_2 is given by:

$$\frac{\partial^2 f}{\partial z_2^2} = \frac{2r_2y_2x_2(2a_2y_2 - z_2)}{z_2}$$

The positive value of the second derivative implies the presence of convexity.

Based on the observed indications of the second derivatives, it may be deduced that the issue does not exhibit global convexity or concavity.

LIST OF REFERENCES

LIST OF REFERENCES

[1] K. Sultan, H. Ali, and Z. Zhang, "Big data perspective and challenges in next generation networks," *Future Internet*, vol. 10, no. 7, p. 56, 2018.

[2] J. Lee, F. Solat, T. Y. Kim, and H. V. Poor, "Federated learning-empowered mobile network management for 5g and beyond networks: From access to core," *IEEE Communications Surveys & Tutorials*, 2024.

[3] L. Banda, M. Mzyece, and F. Mekuria, "5g business models for mobile network operators—a survey," *Ieee Access*, vol. 10, pp. 94851–94886, 2022.

[4] K. N. Qureshi, S. Din, G. Jeon, and F. Piccialli, "Internet of vehicles: Key technologies, network model, solutions and challenges with future aspects," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 3, pp. 1777–1786, 2020.

[5] W. Chen, X. Lin, J. Lee, A. Toskala, S. Sun, C. F. Chiasserini, and L. Liu, "5g-advanced toward 6g: Past, present, and future," *IEEE Journal on Selected Areas in Communications*, vol. 41, no. 6, pp. 1592–1619, 2023.

[6] R. Chataut and R. Akl, "Massive mimo systems for 5g and beyond networks—overview, recent trends, challenges, and future research direction," *Sensors*, vol. 20, no. 10, p. 2753, 2020.

[7] B. E. Shinde and V. Vijayabaskar, "Integrated lte and wi-fi network architecture with authentication of user equipment for dropping off the surplus load of lte," *Wireless Personal Communications*, vol. 125, no. 2, pp. 1469–1481, 2022.

[8] P. Beming, L. Frid, G. Hall, P. Malm, T. Noren, M. Olsson, and G. Rune, "Lte-sae architecture and performance," *Ericsson Review*, vol. 3, pp. 98–104, 2007.

[9] M. Olsson, C. Mulligan, S. Rommer, S. Sultana, and L. Frid, *SAE and the Evolved Packet Core: Driving the mobile broadband revolution*. Academic Press, 2009.

[10] M. Carvalho, E. de Britto, V. F. Silva, D. F. Macedo, *et al.*, "Qd4g: Qoe para vídeo em redes d2d/4g com aprendizado de máquina," in *Anais do XXXVII Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos*, pp. 183–196, SBC, 2019.

[11] S. A. Lonkar and K. V. Reddy, "Analysis of audio and video quality of voice over lte (volte) call," *International Journal of Information Technology*, vol. 14, no. 4, pp. 1981–1994, 2022.

[12] T. Gorman, H. Larijani, *et al.*, "Voice over lte quality evaluation using convolutional neural networks," in *2020 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–7, IEEE, 2020.

[13] T.-T. Tran, Y. Shin, and O.-S. Shin, "Overview of enabling technologies for 3gpp lte-advanced," *EURASIP Journal on Wireless Communications and Networking*, vol. 2012, no. 1, p. 1, 2012.

[14] S. A. Bassam, W. Chen, M. Helaoui, and F. M. Ghannouchi, "Transmitter architecture for ca: Carrier aggregation in lte-advanced systems," *IEEE Microwave Magazine*, vol. 14, no. 5, pp. 78–86, 2013.

[15] S. Mathur, Y. Chaba, and A. Noliya, "Performance analysis of support vector machine learning based carrier aggregation resource scheduling in 5g mobile communication," *Procedia Computer Science*, vol. 218, pp. 2776–2785, 2023.

[16] Y. H. Chiang and W. Liao, "Green multicell cooperation in heterogeneous networks with hybrid energy sources," *IEEE Transactions on Wireless Communications*, vol. 15, pp. 7911–7925, Dec 2016.

[17] R. Irmer, H. Droste, P. Marsch, M. Grieger, G. Fettweis, S. Brueck, H. P. Mayer, L. Thiele, and V. Jungnickel, "Coordinated multipoint: Concepts, performance, and field trial results," *IEEE Communications Magazine*, vol. 49, pp. 102–111, February 2011.

[18] C. Yang, Y. Yao, Z. Chen, and B. Xia, "Analysis on cache-enabled wireless heterogeneous networks," *IEEE Transactions on Wireless Communications*, vol. 15, no. 1, pp. 131–145, 2016.

[19] S. Deb, P. Monogioudis, J. Miernik, and J. P. Seymour, "Algorithms for enhanced inter-cell interference coordination (eicic) in lte hetnets," *IEEE/ACM Transactions on Networking*, vol. 22, pp. 137–150, Feb 2014.

[20] V. Sciancalepore, V. Mancuso, and A. Banchs, "Basics: Scheduling base stations to mitigate interferences in cellular networks," in *World of Wireless, Mobile and Multimedia Networks (WoWMoM), 2013 IEEE 14th International Symposium*, pp. 1–9, June 2013.

[21] M. Z. Chowdhury, M. Shahjalal, S. Ahmed, and Y. M. Jang, "6g wireless communication systems: Applications, requirements, technologies, challenges, and research directions," *IEEE Open Journal of the Communications Society*, vol. 1, pp. 957–975, 2019.

[22] N. Bhushan, J. Li, D. Malladi, R. Gilmore, D. Brenner, A. Damnjanovic, R. T. Sukhavasi, C. Patel, and S. Geirhofer, "Network densification: the dominant theme for wireless evolution into 5g," *IEEE Communications Magazine*, vol. 52, pp. 82–89, February 2014.

[23] M. Moltafet, P. Azmi, and N. Mokari, "Power minimization in 5g heterogeneous cellular networks," in *2016 24th Iranian Conference on Electrical Engineering (ICEE)*, pp. 234–238, May 2016.

[24] A. Gupta and R. K. Jha, "A survey of 5g network: Architecture and emerging technologies," *IEEE Access*, vol. 3, pp. 1206–1232, 2015.

[25] M. Liyanage, A. B. Abro, M. Ylianttila, and A. Gurtov, "Opportunities and challenges of software-defined mobile networks in network security," *IEEE Security Privacy*, vol. 14, pp. 34–44, July 2016.

[26] N. Xiong, W. Han, and A. Vandenberg, "Green cloud computing schemes based on networks: a survey," *IET Communications*, vol. 6, pp. 3294–3300, Dec 2012.

[27] K. Sundaresan, M. Y. Arslan, S. Singh, S. Rangarajan, and S. V. Krishnamurthy, "Fluidnet: A flexible cloud-based radio access network for small cells," *IEEE/ACM Transactions on Networking*, vol. 24, pp. 915–928, April 2016.

[28] X. Huang, G. Xue, R. Yu, and S. Leng, "Joint scheduling and beamforming coordination in cloud radio access networks with qos guarantees," *IEEE Transactions on Vehicular Technology*, vol. 65, pp. 5449–5460, July 2016.

[29] Y. Liao, L. Song, Y. LI, and Y. Zhang, "How much computing capability is enough to run a cloud radio access network?," *IEEE Communications Letters*, vol. PP, no. 99, pp. 1–1, 2016.

[30] B. C. Sahoo, S. R. Mishra, D. Dash, and K. D. Sa, "Design and validation of an antenna array for cloud radio access network applications," in *2020 IEEE 1st International Conference for Convergence in Engineering (ICCE)*, pp. 295–299, IEEE, 2020.

[31] L. Kundu, X. Lin, E. Agostini, and V. Ditya, "Hardware acceleration for open radio access networks: A contemporary overview," *arXiv preprint arXiv:2305.09588*, 2023.

[32] D. Dik and M. S. Berger, "Open-ran fronthaul transport security architecture and implementation," *IEEE Access*, 2023.

[33] S. J. Seelam, S. Andra, and P. C. Jain, "Impact of remote radio head on 5g open-ran technology," in *2022 8th International Conference on Signal Processing and Communication (ICSC)*, pp. 131–136, IEEE, 2022.

[34] M. Polese, L. Bonati, S. D'oro, S. Basagni, and T. Melodia, "Understanding o-ran: Architecture, interfaces, algorithms, security, and research challenges," *IEEE Communications Surveys & Tutorials*, 2023.

[35] K. Ali and M. Jammal, "Traffic forecasting for open radio access networks virtualized network functions in 5g networks," *International Journal of Computer and Information Engineering*, vol. 17, no. 1, pp. 19–27, 2023.

[36] A. Ndao, X. Lagrange, N. Huin, G. Texier, and L. Nuaymi, "Optimal placement of virtualized dus in o-ran architecture," in *2023 IEEE 97th Vehicular Technology Conference (VTC2023-Spring)*, pp. 1–6, IEEE, 2023.

[37] A. Arnaz, J. Lipman, M. Abolhasan, and M. Hiltunen, "Towards integrating intelligence and programmability in open radio access networks: A comprehensive survey," *Ieee Access*, 2022.

[38] A. S. Abdalla, P. S. Upadhyaya, V. K. Shah, and V. Marojevic, "Toward next generation open radio access networks: What o-ran can and cannot do!," *IEEE Network*, vol. 36, no. 6, pp. 206–213, 2022.

[39] S. K. Singh, R. Singh, and B. Kumbhani, "The evolution of radio access network towards open-ran: Challenges and opportunities," in *2020 IEEE Wireless Communications and Networking Conference Workshops (WCNCW)*, pp. 1–6, IEEE, 2020.

[40] E. Coronado, S. N. Khan, and R. Riggio, "5g-empower: A software-defined networking platform for 5g radio access networks," *IEEE Transactions on Network and Service Management*, vol. 16, no. 2, pp. 715–728, 2019.

[41] M. Dryjański, Ł. Kułacz, and A. Kliks, "Toward modular and flexible open ran implementations in 6g networks: Traffic steering use case and o-ran xapps," *Sensors*, vol. 21, no. 24, p. 8173, 2021.

[42] W. Vereecken, W. Van Heddeghem, M. Deruyck, B. Puype, B. Lannoo, W. Joseph, D. Colle, L. Martens, and P. Demeester, "Power consumption in telecommunication networks: overview and reduction strategies," *IEEE Communications Magazine*, vol. 49, no. 6, pp. 62–69, 2011.

[43] J. He, P. Loskot, T. O'Farrell, V. Friderikos, S. Armour, and J. Thompson, "Energy efficient architectures and techniques for green radio access networks," in *2010 5th International ICST Conference on Communications and Networking in China*, pp. 1–6, IEEE, 2010.

[44] A. Khalili Sadaghiani and B. Forouzandeh, "Low-power hardware-efficient memory-based dct processor," *Journal of Real-Time Image Processing*, vol. 19, no. 6, pp. 1105–1121, 2022.

[45] T. Simunic, "Dynamic management of power consumption," in *Power aware computing*, pp. 101–125, Springer, 2002.

[46] Y. A. Sambo, M. Z. Shakir, K. A. Qaraqe, E. Serpedin, and M. A. Imran, "Expanding cellular coverage via cell-edge deployment in heterogeneous networks: Spectral efficiency and backhaul power consumption perspectives," *IEEE Communications Magazine*, vol. 52, no. 6, pp. 140–149, 2014.

[47] V. Bhangdiya, "Low power consumption of led street light based on smart control system," in *2016 International Conference on Global Trends in Signal Processing, Information Computing and Communication (ICGTSPICC)*, pp. 619–622, IEEE, 2016.

[48] J. Oliver, O. Mocanu, and C. Ferrer, "Energy awareness through software optimisation as a performance estimate case study of the mc68hc908gp32 microcontroller," in *Proceedings. 4th International Workshop on Microprocessor Test and Verification-Common Challenges and Solutions*, pp. 111–116, IEEE, 2003.

[49] S. M. Ghafari, M. Fazeli, A. Patooghy, and L. Rikhtechi, "Bee-mmt: A load balancing method for power consumption management in cloud computing," in *2013 Sixth International Conference on Contemporary Computing (IC3)*, pp. 76–80, IEEE, 2013.

[50] A. Qazi, F. Hussain, N. A. Rahim, G. Hardaker, D. Alghazzawi, K. Shaban, and K. Haruna, "Towards sustainable energy: a systematic review of renewable energy sources, technologies, and public opinions," *IEEE access*, vol. 7, pp. 63837–63851, 2019.

[51] D. Sharma, S. Singhal, A. Rai, and A. Singh, "Analysis of power consumption in standalone 5g network and enhancement in energy efficiency using a novel routing protocol," *Sustainable Energy, Grids and Networks*, vol. 26, p. 100427, 2021.

[52] T. Pamuklu, S. Mollahasani, and M. Erol-Kantarci, "Energy-efficient and delay-guaranteed joint resource allocation and du selection in o-ran," in *2021 IEEE 4th 5G World Forum (5GWF)*, pp. 99–104, IEEE, 2021.

[53] M. Dayarathna, Y. Wen, and R. Fan, "Data center energy consumption modeling: A survey," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 1, pp. 732–794, 2015.

[54] R. S. Alhumaima and H. S. Al-Raweshidy, "Modelling the power consumption and trade-offs of virtualised cloud radio access networks," *IET Communications*, vol. 11, no. 7, pp. 1158–1164, 2017.

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

[56] H. Holtkamp, G. Auer, V. Giannini, and H. Haas, "A parameterized base station power model," *IEEE Communications Letters*, vol. 17, no. 11, pp. 2033–2035, 2013.

[57] R. S. Alhumaima, M. Khan, and H. S. Al-Raweshidy, "Component and parameterised power model for cloud radio access network," *IET Communications*, vol. 10, no. 7, pp. 745–752, 2016.

[58] R. S. Alhumaima and H. S. Al-Raweshidy, "Evaluating the energy efficiency of software defined-based cloud radio access networks," *IET Communications*, vol. 10, no. 8, pp. 987–994, 2016.

[59] R. Shea, H. Wang, and J. Liu, "Power consumption of virtual machines with network transactions: Measurement and improvements," in *IEEE INFOCOM 2014-IEEE Conference on Computer Communications*, pp. 1051–1059, IEEE, 2014.

[60] Q. Huang, F. Gao, R. Wang, and Z. Qi, "Power consumption of virtual machine live migration in clouds," in *2011 third international conference on communications and mobile computing*, pp. 122–125, IEEE, 2011.

[61] M. Marcu and D. Tudor, "Power consumption measurements of virtual machines," in *2011 6th IEEE International Symposium on Applied Computational Intelligence and Informatics (SACI)*, pp. 445–449, IEEE, 2011.

[62] R. S. Alhumaima, R. K. Ahmed, and H. S. Al-Raweshidy, "Maximizing the energy efficiency of virtualized c-ran via optimizing the number of virtual machines," *IEEE Transactions on Green Communications and Networking*, vol. 2, no. 4, pp. 992–1001, 2018.

[63] Y. Al-Karawi, R. S. Alhumaima, and H. Al-Raweshidy, “Quality of service of quantum entanglement in mobile networks,” *IEEE Access*, vol. 9, pp. 167242–167251, 2021.

[64] K. Fang, J. Zhao, X. Li, Y. Li, and R. Duan, “Quantum network: from theory to practice,” *Science China Information Sciences*, vol. 66, no. 8, p. 180509, 2023.

[65] S. Muralidharan, L. Li, J. Kim, N. Lütkenhaus, M. D. Lukin, and L. Jiang, “Optimal architectures for long distance quantum communication,” *Scientific reports*, vol. 6, no. 1, p. 20463, 2016.

[66] J. Miguel-Ramiro, A. Pirker, and W. Dür, “Optimized quantum networks,” *Quantum*, vol. 7, p. 919, 2023.

[67] I. Šupić, J. Bowles, M.-O. Renou, A. Acín, and M. J. Hoban, “Quantum networks self-test all entangled states,” *Nature Physics*, pp. 1–6, 2023.

[68] S. Subramani and S. K. Svn, “Review of security methods based on classical cryptography and quantum cryptography,” *Cybernetics and Systems*, pp. 1–19, 2023.

[69] Z. Qu, Z. Chen, X. Ning, and P. Tiwari, “Qepp: A quantum efficient privacy protection protocol in 6g-quantum internet of vehicles,” *IEEE Transactions on Intelligent Vehicles*, 2023.

[70] J. Choi, S. Oh, and J. Kim, “Energy-efficient cluster head selection via quantum approximate optimization,” *Electronics*, vol. 9, no. 10, p. 1669, 2020.

[71] M. Erhard, M. Krenn, and A. Zeilinger, “Advances in high-dimensional quantum entanglement,” *Nature Reviews Physics*, vol. 2, no. 7, pp. 365–381, 2020.

[72] H.-J. Briegel, W. Dür, J. I. Cirac, and P. Zoller, “Quantum repeaters: the role of imperfect local operations in quantum communication,” *Physical Review Letters*, vol. 81, no. 26, p. 5932, 1998.

[73] M. Aspelmeyer, T. Jennewein, M. Pfennigbauer, W. R. Leeb, and A. Zeilinger, “Long-distance quantum communication with entangled photons using satellites,” *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 9, no. 6, pp. 1541–1551, 2003.

[74] L. Gyongyosi and S. Imre, “Distributed quantum computation for near-term quantum environments,” in *Quantum Information Science, Sensing, and Computation XIII*, vol. 11726, p. 117260I, International Society for Optics and Photonics, 2021.

[75] I. P. Christov, “Spatial entanglement of fermions in one-dimensional quantum dots,” *Entropy*, vol. 23, no. 7, p. 868, 2021.

[76] M. Chehimi and W. Saad, “Entanglement rate optimization in heterogeneous quantum communication networks,” *arXiv preprint arXiv:2105.14507*, 2021.

[77] K. Boström and T. Felbinger, “Deterministic secure direct communication using entanglement,” *Physical Review Letters*, vol. 89, no. 18, p. 187902, 2002.

[78] C.-Z. Peng, T. Yang, X.-H. Bao, J. Zhang, X.-M. Jin, F.-Y. Feng, B. Yang, J. Yang, J. Yin, Q. Zhang, *et al.*, "Experimental free-space distribution of entangled photon pairs over 13 km: towards satellite-based global quantum communication," *Physical review letters*, vol. 94, no. 15, p. 150501, 2005.

[79] M.-H. Hsieh and M. M. Wilde, "Entanglement-assisted communication of classical and quantum information," *IEEE Transactions on Information Theory*, vol. 56, no. 9, pp. 4682–4704, 2010.

[80] K. Banaszek, A. Dragan, W. Wasilewski, and C. Radzewicz, "Experimental demonstration of entanglement-enhanced classical communication over a quantum channel with correlated noise," *Physical review letters*, vol. 92, no. 25, p. 257901, 2004.

[81] N. Min, J. Jin-ya, and L. Xiao-hui, "A novel optimum quantum states entanglement multiplexing and relay scheme for land quantum mobile communication," *Acta Photonica Sinica*, vol. 40, no. 5, p. 774, 2011.

[82] S. D. Bartlett, T. Rudolph, and R. W. Spekkens, "Classical and quantum communication without a shared reference frame," *Physical review letters*, vol. 91, no. 2, p. 027901, 2003.

[83] B. Da Lio, D. Bacco, D. Cozzolino, N. Biagi, T. N. Arge, E. Larsen, K. Rottwitt, Y. Ding, A. Zavatta, and L. K. Oxenløwe, "Stable transmission of high-dimensional quantum states over a 2-km multicore fiber," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 26, no. 4, pp. 1–8, 2019.

[84] Y. Yang, G. Chiribella, and M. Hayashi, "Communication cost of quantum processes," *IEEE Journal on Selected Areas in Information Theory*, vol. 1, no. 2, pp. 387–400, 2020.

[85] J. Rubio, P. A. Knott, T. J. Proctor, and J. A. Dunningham, "Quantum sensing networks for the estimation of linear functions," *Journal of Physics A: Mathematical and Theoretical*, vol. 53, no. 34, p. 344001, 2020.

[86] J. S. Sidhu, Y. Ouyang, E. T. Campbell, and P. Kok, "Tight bounds on the simultaneous estimation of incompatible parameters," *Physical Review X*, vol. 11, no. 1, p. 011028, 2021.

[87] R. Jozsa, D. S. Abrams, J. P. Dowling, and C. P. Williams, "Quantum clock synchronization based on shared prior entanglement," *Physical Review Letters*, vol. 85, no. 9, p. 2010, 2000.

[88] J. S. Sidhu, S. Izumi, J. S. Neergaard-Nielsen, C. Lupo, and U. L. Andersen, "Quantum receiver for phase-shift keying at the single-photon level," *Prx Quantum*, vol. 2, no. 1, p. 010332, 2021.

[89] S. Pirandola, U. L. Andersen, L. Banchi, M. Berta, D. Bunandar, R. Colbeck, D. Englund, T. Gehring, C. Lupo, C. Ottaviani, *et al.*, "Advances in quantum cryptography," *Advances in optics and photonics*, vol. 12, no. 4, pp. 1012–1236, 2020.

[90] E. Knill, R. Laflamme, and G. J. Milburn, "A scheme for efficient quantum computation with linear optics," *nature*, vol. 409, no. 6816, pp. 46–52, 2001.

[91] D. Braun, G. Adesso, F. Benatti, R. Floreanini, U. Marzolino, M. W. Mitchell, and S. Pirandola, “Quantum-enhanced measurements without entanglement,” *Reviews of Modern Physics*, vol. 90, no. 3, p. 035006, 2018.

[92] C. Schimpf, M. Reindl, F. Basso Basset, K. D. Jöns, R. Trotta, and A. Rastelli, “Quantum dots as potential sources of strongly entangled photons: Perspectives and challenges for applications in quantum networks,” *Applied Physics Letters*, vol. 118, no. 10, 2021.

[93] M. Zukowski, A. Zeilinger, M. A. Horne, and A. K. Ekert, “" event-ready-detectors" bell experiment via entanglement swapping.,” *Physical review letters*, vol. 71, no. 26, 1993.

[94] D. Deutsch, A. Ekert, R. Jozsa, C. Macchiavello, S. Popescu, and A. Sanpera, “Quantum privacy amplification and the security of quantum cryptography over noisy channels,” *Physical review letters*, vol. 77, no. 13, p. 2818, 1996.

[95] S. Pirandola, “End-to-end capacities of a quantum communication network,” *Communications Physics*, vol. 2, no. 1, p. 51, 2019.

[96] E. O. Ilo-Okeke, L. Tessler, J. P. Dowling, and T. Byrnes, “Remote quantum clock synchronization without synchronized clocks,” *npj Quantum Information*, vol. 4, no. 1, p. 40, 2018.

[97] S. Wehner, D. Elkouss, and R. Hanson, “Quantum internet: A vision for the road ahead,” *Science*, vol. 362, no. 6412, p. eaam9288, 2018.

[98] M. Zwerger, A. Pirker, V. Dunjko, H. J. Briegel, and W. Dür, “Long-range big quantum-data transmission,” *Physical review letters*, vol. 120, no. 3, p. 030503, 2018.

[99] Q. Zhang, F. Xu, L. Li, N.-L. Liu, and J.-W. Pan, “Quantum information research in china,” *Quantum Science and Technology*, vol. 4, no. 4, p. 040503, 2019.

[100] S. Santra and V. S. Malinovsky, “Quantum networking with short-range entanglement assistance,” *Physical Review A*, vol. 103, no. 1, p. 012407, 2021.

[101] M. Bérces and S. Imre, “Extension and analysis of modified superdense-coding in multi-user environment,” in *2015 IEEE 19th International Conference on Intelligent Engineering Systems (INES)*, pp. 291–294, IEEE, 2015.

[102] I. B. Djordjevic, “Integrated optics modules based proposal for quantum information processing, teleportation, qkd, and quantum error correction employing photon angular momentum,” *IEEE Photonics Journal*, vol. 8, no. 1, pp. 1–12, 2016.

[103] M. A. Nielsen, “Conditions for a class of entanglement transformations,” *Physical Review Letters*, vol. 83, no. 2, p. 436, 1999.

[104] S. Imre and F. Balazs, *Quantum computing and communications*. Wiley Online Library, 2005.

[105] B. Schumacher and M. A. Nielsen, “Quantum data processing and error correction,” *Physical Review A*, vol. 54, no. 4, p. 2629, 1996.

[106] M. Hayashi and K. Matsumoto, "Variable length universal entanglement concentration by local operations and its application to teleportation and dense coding," *arXiv preprint quant-ph/0109028*, 2001.

[107] J. Li, J. Xiong, Q. Zhang, L. Zhong, Y. Zhou, J. Li, and X. Lu, "A one-time pad encryption method combining full-phase image encryption and hiding," *Journal Of Optics*, vol. 19, no. 8, p. 085701, 2017.

[108] J. G. Rarity, P. Owens, and P. Tapster, "Quantum random-number generation and key sharing," *Journal of Modern Optics*, vol. 41, no. 12, pp. 2435–2444, 1994.

[109] V. Giovannetti, A. S. Holevo, S. Lloyd, and L. Maccone, "Generalized minimal output entropy conjecture for one-mode gaussian channels: definitions and some exact results," *Journal of Physics A: Mathematical and Theoretical*, vol. 43, no. 41, p. 415305, 2010.

[110] D. Liu, L. Wang, Y. Chen, M. Elkashlan, K.-K. Wong, R. Schober, and L. Hanzo, "User association in 5g networks: A survey and an outlook," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 2, pp. 1018–1044, 2016.

[111] M. Marcus and B. Pattan, "Millimeter wave propagation: spectrum management implications," *IEEE Microwave Magazine*, vol. 6, no. 2, pp. 54–62, 2005.

[112] H. Hantouti, N. Benamar, T. Taleb, and A. Laghrissi, "Traffic steering for service function chaining," *IEEE Communications Surveys & Tutorials*, vol. 21, no. 1, pp. 487–507, 2018.

[113] Z. Xiao, W. Song, and Q. Chen, "Dynamic resource allocation using virtual machines for cloud computing environment," *IEEE transactions on parallel and distributed systems*, vol. 24, no. 6, pp. 1107–1117, 2012.

[114] W.-Z. Zhang, I. A. Elgendy, M. Hammad, A. M. Iliyasu, X. Du, M. Guizani, and A. A. Abd El-Latif, "Secure and optimized load balancing for multitier iot and edge-cloud computing systems," *IEEE Internet of Things Journal*, vol. 8, no. 10, pp. 8119–8132, 2020.

[115] R. A. Novo, C. J. Davolos, and Z. J. Zhao, "Measuring the impact of redirecting and offloading mobile data traffic," *Bell Labs Technical Journal*, vol. 18, no. 1, pp. 81–103, 2013.

[116] A. H. Alhilali and A. Montazerolghaem, "Artificial intelligence based load balancing in sdn: A comprehensive survey," *Internet of Things*, p. 100814, 2023.

[117] S. Zhang, "An overview of network slicing for 5g," *IEEE Wireless Communications*, vol. 26, no. 3, pp. 111–117, 2019.

[118] D. Attanayaka, P. Porambage, M. Liyanage, and M. Ylianttila, "Peer-to-peer federated learning based anomaly detection for open radio access networks,"

[119] P. K. Thiruvasagam, V. Venkataram, V. R. Ilangovan, M. Perapalla, R. Payyanur, V. Kumar, *et al.*, "Open ran: Evolution of architecture, deployment aspects, and future directions," *arXiv preprint arXiv:2301.06713*, 2023.

[120] I. Oussakel, P. Owezarski, P. Berthou, and L. Houssin, "Toward radio access network slicing enforcement in multi-cell 5g system," *Journal of Network and Systems Management*, vol. 31, no. 1, p. 8, 2023.

[121] S. Mondal and M. Ruffini, "Fairness guaranteed and auction-based x-haul and cloud resource allocation in multi-tenant o-rans," *arXiv preprint arXiv:2301.00597*, 2023.

[122] H.-M. Yoo, J.-S. Rhee, S.-Y. Bang, and E.-K. Hong, "Load balancing algorithm running on open ran ric," in *2022 13th International Conference on Information and Communication Technology Convergence (ICTC)*, pp. 1226–1228, 2022.

[123] M. Peng and K. Zhang, "Recent advances in fog radio access networks: Performance analysis and radio resource allocation," *IEEE Access*, vol. 4, pp. 5003–5009, 2016.

[124] Y. Ye, T. Zhang, and L. Yang, "Joint user association and resource allocation for load balancing in ran slicing," *Physical Communication*, vol. 49, p. 101459, 2021.

[125] E. Amiri, N. Wang, M. Shojafar, and R. Tafazolli, "Optimizing virtual network function splitting in open-ran environments," in *2022 IEEE 47th Conference on Local Computer Networks (LCN)*, pp. 422–429, IEEE, 2022.

[126] C.-H. Lai, L.-H. Shen, and K.-T. Feng, "Intelligent load balancing and resource allocation in o-ran: A multi-agent multi-armed bandit approach," *arXiv preprint arXiv:2303.14355*, 2023.

[127] O. Orhan, V. N. Swamy, T. Tetzlaff, M. Nassar, H. Nikopour, and S. Talwar, "Connection management xapp for o-ran ric: A graph neural network and reinforcement learning approach," in *2021 20th IEEE International Conference on Machine Learning and Applications (ICMLA)*, pp. 936–941, IEEE, 2021.

[128] N. Budhdev, A. Maity, M. C. Chan, and T. Mitra, "Load balancing for a user-level virtualized 5g cloud-ran," in *Proceedings of the 17th ACM Workshop on Mobility in the Evolving Internet Architecture*, pp. 1–6, 2022.

[129] E. A. R. da Paixão, R. F. Vieira, W. V. Araújo, and D. L. Cardoso, "Optimized load balancing by dynamic bbu-rrh mapping in c-ran architecture," in *2018 Third International Conference on Fog and Mobile Edge Computing (FMEC)*, pp. 100–104, IEEE, 2018.

[130] B. Mahapatra, R. Kumar, A. Turuk, and S. Patra, "Clb: a multilevel co-operative load balancing algorithm for c-ran architecture," *Digital Communications and Networks*, vol. 5, no. 4, pp. 308–316, 2019.

[131] B. Brik, K. Boutiba, and A. Ksentini, "Deep learning for b5g open radio access network: Evolution, survey, case studies, and challenges," *IEEE Open Journal of the Communications Society*, vol. 3, pp. 228–250, 2022.

[132] S. Niknam, A. Roy, H. S. Dhillon, S. Singh, R. Banerji, J. H. Reed, N. Saxena, and S. Yoon, "Intelligent o-ran for beyond 5g and 6g wireless networks," *arXiv preprint arXiv:2005.08374*, 2020.

[133] B. Balasubramanian, E. S. Daniels, M. Hiltunen, R. Jana, K. Joshi, R. Sivaraj, T. X. Tran, and C. Wang, "Ric: A ran intelligent controller platform for ai-enabled cellular networks," *IEEE Internet Computing*, vol. 25, no. 2, pp. 7–17, 2021.

[134] N. Anerousis, P. Chemouil, A. A. Lazar, N. Mihai, and S. B. Weinstein, "The origin and evolution of open programmable networks and sdn," *IEEE Communications Surveys & Tutorials*, vol. 23, no. 3, pp. 1956–1971, 2021.

[135] O. Alliance, "O-ran use cases and deployment scenarios whitepaper," *O-RAN ALLIANCE, Tech. Rep*, 2020.

[136] T. Senevirathna, Z. Salazar, V. H. La, S. Marchal, B. Siniarski, M. Liyanage, and S. Wang, "A survey on xai for beyond 5g security: Technical aspects, use cases, challenges and research directions," *arXiv preprint arXiv:2204.12822*, 2022.

[137] R. Doriguzzi-Corin, S. Scott-Hayward, D. Siracusa, M. Savi, and E. Salvadori, "Dynamic and application-aware provisioning of chained virtual security network functions," *IEEE Transactions on Network and Service Management*, vol. 17, no. 1, pp. 294–307, 2019.

[138] E. Zeydan, J. Mangues-Bafalluy, J. Baranda, M. Requena, and Y. Turk, "Service based virtual ran architecture for next generation cellular systems," *IEEE Access*, vol. 10, pp. 9455–9470, 2022.

[139] K. Samdanis and T. Taleb, "The road beyond 5g: A vision and insight of the key technologies," *IEEE Network*, vol. 34, no. 2, pp. 135–141, 2020.

[140] M. Polese, L. Bonati, S. D'Oro, S. Basagni, and T. Melodia, "Understanding o-ran: Architecture, interfaces, algorithms, security, and research challenges," *arXiv preprint arXiv:2202.01032*, 2022.

[141] A. Vlahov, D. Ekova, V. Poulkov, and T. Cooklev, "Virtualized, open and intelligent: The evolution of the radio access network," in *6G Enabling Technologies*, pp. 181–214, River Publishers, 2023.

[142] H. K. Sabat, "The evolving mobile wireless value chain and market structure," *Telecommunications policy*, vol. 26, no. 9-10, pp. 505–535, 2002.

[143] M. Iwamura, K. Etemad, M.-H. Fong, R. Nory, and R. Love, "Carrier aggregation framework in 3gpp lte-advanced [wimax/lte update]," *IEEE Communications magazine*, vol. 48, no. 8, pp. 60–67, 2010.

[144] M. Filo, J. D. L. Ducoing, C. Jayawardena, C. Husmann, R. Tafazolli, and K. Nikitopoulos, "Evaluating non-linear beamforming in a 3gpp-compliant framework using the sword platform," in *2020 IEEE 31st Annual International Symposium on Personal, Indoor and Mobile Radio Communications*, pp. 1–6, IEEE, 2020.

[145] A. Ghosh, A. Maeder, M. Baker, and D. Chandramouli, "5g evolution: A view on 5g cellular technology beyond 3gpp release 15," *IEEE access*, vol. 7, pp. 127639–127651, 2019.

[146] Y. Li, T. Jiang, K. Luo, and S. Mao, "Green heterogeneous cloud radio access networks: Potential techniques, performance trade-offs, and challenges," *IEEE Communications Magazine*, vol. 55, no. 11, pp. 33–39, 2017.

[147] B. M. Al-Nedawe, R. S. Alhumaima, and W. H. Ali, "On the quality of service of next generation green networks," *IET Networks*, vol. 11, no. 1, pp. 1–12, 2022.

[148] C.-H. Lin, W.-C. Chien, J.-Y. Chen, C.-F. Lai, and H.-C. Chao, "Energy efficient fog ran (f-ran) with flexible bbu resource assignment for latency aware mobile edge computing (mec) services," in *2019 IEEE 90th Vehicular Technology Conference (VTC2019-Fall)*, pp. 1–6, IEEE, 2019.

[149] M. Deruyck, W. Joseph, and L. Martens, "Power consumption model for macrocell and microcell base stations," *Transactions on Emerging Telecommunications Technologies*, vol. 25, no. 3, pp. 320–333, 2014.

[150] K. Martsenko, "Developing and evaluating power models of heterogeneous computer systems," Master's thesis, ETH Zurich, 2021.

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

[152] K. Liu, J. He, J. Ding, Y. Zhu, and Z. Liu, "Base station power model and application for energy efficient lte," in *2013 15th IEEE International Conference on Communication Technology*, pp. 86–92, IEEE, 2013.

[153] S. Bassoy, M. Jaber, M. A. Imran, and P. Xiao, "Load aware self-organising user-centric dynamic comp clustering for 5g networks," *IEEE Access*, vol. 4, pp. 2895–2906, 2016.

[154] X. Yang, Y. Wang, D. Zhang, and L. Cuthbert, "Resource allocation in lte ofdma systems using genetic algorithm and semi-smart antennas," in *2010 IEEE Wireless Communication and Networking Conference*, pp. 1–6, IEEE, 2010.

[155] M. Peng, K. Zhang, J. Jiang, J. Wang, and W. Wang, "Energy-efficient resource assignment and power allocation in heterogeneous cloud radio access networks," *IEEE Transactions on Vehicular Technology*, vol. 64, no. 11, pp. 5275–5287, 2014.

[156] M. Kim, S. Kasi, P. A. Lott, D. Venturelli, J. Kaewell, and K. Jamieson, "Heuristic quantum optimization for 6g wireless communications," *IEEE Network*, vol. 35, no. 4, pp. 8–15, 2021.

[157] Y.-W. Cho, G. Campbell, J. Everett, J. Bernu, D. Higginbottom, M. Cao, J. Geng, N. Robins, P. Lam, and B. Buchler, "Highly efficient and long-lived optical quantum memory with cold atoms," in *2017 Conference on Lasers and Electro-Optics (CLEO)*, pp. 1–2, IEEE, 2017.

[158] S. Wijesekera, X. Huang, and D. Sharma, "Multi-agent based approach for quantum key distribution in wifi networks," in *KES International Symposium on Agent and Multi-Agent Systems: Technologies and Applications*, pp. 293–303, Springer, 2009.

[159] T. Nagata, R. Okamoto, J. L. O'brien, K. Sasaki, and S. Takeuchi, "Beating the standard quantum limit with four-entangled photons," *Science*, vol. 316, no. 5825, pp. 726–729, 2007.

[160] J. M. Gambetta, J. M. Chow, and M. Steffen, "Building logical qubits in a superconducting quantum computing system," *npj Quantum Information*, vol. 3, no. 1, pp. 1–7, 2017.

[161] B. Fedrici, L. Ngah, O. Alibart, F. Kaiser, L. Labonté, V. D'Auria, and S. Tanzilli, "All-optical synchronization for quantum communication networks," in *2017 19th International Conference on Transparent Optical Networks (ICTON)*, pp. 1–3, IEEE, 2017.

[162] A. Chinnappan and R. Balasubramanian, "Complexity–consistency trade-off in multi-attribute decision making for vertical handover in heterogeneous wireless networks," *IET Networks*, vol. 5, no. 1, pp. 13–21, 2016.

[163] R. S. Alhumaima, S. T. Alwan, and R. K. Ahmed, "Mitigating x2-ap interface cost using quantum teleportation," *IET Networks*, vol. 9, no. 5, pp. 247–254, 2020.

[164] S. Agrawal and D. Lin, *Advances in Cryptology–ASIACRYPT 2022: 28th International Conference on the Theory and Application of Cryptology and Information Security, Taipei, Taiwan, December 5–9, 2022, Proceedings, Part IV*, vol. 13794. Springer Nature, 2023.

[165] B. Rawat, N. Mehra, A. S. Bist, M. Yusup, and Y. P. A. Sanjaya, "Quantum computing and ai: Impacts & possibilities," *ADI Journal on Recent Innovation*, vol. 3, no. 2, pp. 202–207, 2022.

[166] C. Portmann and R. Renner, "Security in quantum cryptography," *Reviews of Modern Physics*, vol. 94, no. 2, p. 025008, 2022.

[167] A. J. Daley, I. Bloch, C. Kokail, S. Flannigan, N. Pearson, M. Troyer, and P. Zoller, "Practical quantum advantage in quantum simulation," *Nature*, vol. 607, no. 7920, pp. 667–676, 2022.

[168] S. Pathak, A. Mani, M. Sharma, and A. Chatterjee, "Decomposition based quantum inspired salp swarm algorithm for multiobjective optimization," *IEEE Access*, vol. 10, pp. 105421–105436, 2022.

[169] X. Wang, Z. Lin, F. Lin, and P. Xiao, "Quantum sensing based joint 3d beam training for uav-mounted star-ris aided terahertz multi-user massive mimo systems," *arXiv preprint arXiv:2212.07731*, 2022.

[170] A. Quirce and A. Valle, "Random polarization switching in gain-switched vcsels for quantum random number generation," *Optics Express*, vol. 30, no. 7, pp. 10513–10527, 2022.

[171] E. Manfreda-Schulz, J. D. Elliot, M. van Niekerk, C. C. Tison, M. L. Fanto, S. F. Preble, and G. A. Howland, "Generation of high-dimensional entanglement on a silicon photonic chip," in *Quantum 2.0*, pp. QTu4B–4, Optica Publishing Group, 2022.

[172] M. Tayyab, G. P. Koudouridis, X. Gelabert, and R. Jäntti, "Signaling overhead and power consumption during handover in lte," in *2019 IEEE Wireless Communications and Networking Conference (WCNC)*, pp. 1–6, IEEE, 2019.

[173] X. Dong and H. Wang, "Enhanced corrections near holographic entanglement transitions: a chaotic case study," *Journal of High Energy Physics*, vol. 2020, no. 11, pp. 1–32, 2020.

[174] P. Calabrese and J. Cardy, "Entanglement entropy and quantum field theory," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2004, no. 06, p. P06002, 2004.

[175] D. Mimran, R. Bitton, Y. Kfir, E. Klevansky, O. Brodt, H. Lehmann, Y. Elovici, and A. Shabtai, "Security of open radio access networks," *Computers & Security*, vol. 122, p. 102890, 2022.

[176] N. Li, G. Liu, H. Zhang, Q. Zhao, Y. Zhao, Z. Tong, Y. Wang, and J. Sun, "Micro-service-based radio access network," *China Communications*, vol. 19, no. 3, pp. 1–15, 2022.

[177] D. Mimran, R. Bitton, Y. Kfir, E. Klevansky, O. Brodt, H. Lehmann, Y. Elovici, and A. Shabtai, "Evaluating the security of open radio access networks," *arXiv preprint arXiv:2201.06080*, 2022.

[178] P. K. Gupta, R. Rajakumar, and C. Kumar, "Energy impact of signalling protocols in 3gpp-lte and guidelines for savings," in *2012 Annual IEEE India Conference (INDICON)*, pp. 126–130, IEEE, 2012.

[179] K. Alexandris, N. Nikaein, R. Knopp, and C. Bonnet, "Analyzing x2 handover in lte/lte-a," in *2016 14th international symposium on modeling and optimization in mobile, ad hoc, and wireless networks (WiOpt)*, pp. 1–7, IEEE, 2016.

[180] M. Shahjalal, W. Kim, W. Khalid, S. Moon, M. Khan, S. Liu, S. Lim, E. Kim, D.-W. Yun, J. Lee, *et al.*, "Enabling technologies for ai empowered 6g massive radio access networks," *ICT Express*, vol. 9, no. 3, pp. 341–355, 2023.

[181] I. Rahman, S. M. Razavi, O. Liberg, C. Hoymann, H. Wiemann, C. Tidestav, P. Schliwa-Bertling, P. Persson, and D. Gerstenberger, "5g evolution toward 5g advanced: An overview of 3gpp releases 17 and 18," *Ericsson Technology Review*, vol. 2021, no. 14, pp. 2–12, 2021.

[182] J. F. Dynes, W. W. Tam, A. Plews, B. Fröhlich, A. W. Sharpe, M. Lucamarini, Z. Yuan, C. Radig, A. Straw, T. Edwards, *et al.*, "Ultra-high bandwidth quantum secured data transmission," *Scientific reports*, vol. 6, no. 1, p. 35149, 2016.

[183] T. Yang, Q. Zhang, T.-Y. Chen, S. Lu, J. Yin, J.-W. Pan, Z.-Y. Wei, J.-R. Tian, and J. Zhang, "Experimental synchronization of independent entangled photon sources," *Physical review letters*, vol. 96, no. 11, p. 110501, 2006.

[184] A. Salmanogli, D. Gokcen, and H. S. Gecim, "Entanglement sustainability in quantum radar," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 26, no. 6, pp. 1–11, 2020.

[185] Y. Chu, X. Zhang, B. Chen, J. Wang, J. Yang, R. Jiang, and M. Hu, "Picosecond high-power 213-nm deep-ultraviolet laser generation using β-bab2o4 crystal," *Optics & Laser Technology*, vol. 134, p. 106657, 2021.

[186] P. Blasiak, E. Borsuk, and M. Markiewicz, "On safe post-selection for bell tests with ideal detectors: Causal diagram approach," *Quantum*, vol. 5, p. 575, 2021.

[187] J. Bergli, Y. M. Galperin, and B. Altshuler, "Decoherence in qubits due to low-frequency noise," *New Journal of Physics*, vol. 11, no. 2, p. 025002, 2009.

[188] A. Fiasconaro, J. J. Mazo, and B. Spagnolo, "Noise-induced enhancement of stability in a metastable system with damping," *Physical Review E*, vol. 82, no. 4, p. 041120, 2010.

[189] L. Funcke, T. Hartung, K. Jansen, S. Kühn, P. Stornati, and X. Wang, "Measurement error mitigation in quantum computers through classical bit-flip correction," *Physical Review A*, vol. 105, no. 6, p. 062404, 2022.

[190] F. Chapeau-Blondeau and E. Belin, "Fourier-transform quantum phase estimation with quantum phase noise," *Signal Processing*, vol. 170, p. 107441, 2020.

[191] B. Yurke and M. Potasek, "Obtainment of thermal noise from a pure quantum state," *Physical Review A*, vol. 36, no. 7, p. 3464, 1987.

[192] M. Hendrych, X. Shi, A. Valencia, and J. P. Torres, "Broadening the bandwidth of entangled photons: A step towards the generation of extremely short biphotons," *Physical Review A*, vol. 79, no. 2, p. 023817, 2009.

[193] C. Pang, A. Hindle, B. Adams, and A. E. Hassan, "What do programmers know about software energy consumption?," *IEEE Software*, vol. 33, no. 3, pp. 83–89, 2015.

[194] M. Plenio, J. Hartley, and J. Eisert, "Dynamics and manipulation of entanglement in coupled harmonic systems with many degrees of freedom," *New Journal of Physics*, vol. 6, no. 1, p. 36, 2004.

[195] Y. Al-Karawi, H. Al-Raweshidy, and R. Nilavalan, "Power consumption evaluation of next generation open radio access network," in *2024 IEEE International Conference on Consumer Electronics (ICCE)*, pp. 1–6, IEEE, 2024.

[196] Y. Al-Karawi, H. Al-Raweshidy, and R. Nilavalan, "Optimizing the energy efficiency using quantum based load balancing in open radio access networks," *IEEE Access*, 2024.

list of publications

1. Al-Karawi, Yassir, Raad S. Alhumaima, and Hamed S. Al-Raweshidy. "Quality of Service of Quantum Entanglement in Mobile Networks." *IEEE Access* 9 (2021): 167242-167251 [63].
2. Al-Karawi, Yassir, Hamed Al-Raweshidy, and Rajagopal Nilavalan. "Power Consumption Evaluation of Next Generation Open Radio Access Network."*2024 IEEE International Conference on Consumer Electronics (ICCE)*. 2024 [195].
3. Y. Al-Karawi, H. Al-Raweshidy and R. Nilavalan, "Optimizing the Energy Efficiency Using Quantum Based Load Balancing in Open Radio Access Networks," *IEEE Access*, vol. 12, pp. 37903-37918, (2024) [196].
4. Al-Karawi, Yassir, and Hamed S. Al-Raweshidy."Quantum-Enabled Method for Power, Delay and Energy Efficiency Enhancement in Open Radio Access Networks.",*IEEE Transactions on Green Communications and Networking*, submitted.