Architectural Considerations and Resource Allocation In Energy Efficient Networking

Mlungisi Molefe

Submitted to the Faculty of Engineering and the Built Environment

in Partial Fulfilment of the Requirements for the

Master of Engineering in Electronic Engineering

September 2022

Approved for Final Submission

Supervisor: Professor B Nleya Student: Mlungisi Molefe

Date____28/09/2022_____ Date ____27/09/2022___

Acknowledgement

My sincere thanks to my family and friends for their unconditional as well as unconditioned support throughout pursuing this work. In particular, I am highly indebted to my family. I appreciate the spiritual support of my family. My mother and family members have given me unequivocal support throughout my studies and have always been with me despite the distance. Further, appreciating the unconditional guidance from the academic supervisor Prof. Nleya, I am truly grateful for the opportunity given to me, above all, I am privileged to have worked with such a great supervisor. I also extend my gratitude to colleagues in the Postgraduate Laboratory for constantly encouraging me throughout my research. Finally, I would also like to extend my sincere gratitude to God, for this work would not be possible without his willingness.

.

Table of Contents

Plagiarism Declaration

1. I know and understand that plagiarism is using another person's work and pretending it is one's own, which is wrong.

2. This report is my own work.

3. I have appropriately referenced the work of other people I have used.

4. I have not allowed, and will not allow, anyone to copy my work with the intention of passing it off as his/her own work.

Surname and Initials	Student Number
MM Molefe	21306596

Signature:

Date: 30/05/2022

Abstract

Word-wide data traffic is continuously surging, triggered mainly by the emergence of Internet-of-Things (IoT)'s services and Fog-Cloud computing-based applications. This calls for existing optical and wireless-based network infrastructures to upgrade capacity accordingly to meet required massive bandwidth demands to accommodate the ever-surging data traffic volumes. However, continuously elevating the resource requirements in terms of bandwidth provisioning implies increasing the number of energy-consuming network elements, which will increase overall operational expenditures and carbon footprint due to extra power generation. Carbon emissions contribute significantly to global warming. To avert this, it has become necessary to promote energy-efficient networking. For that reason, it necessitated an emphasis on energy efficiency in the design, operation, and planning of transport networks.

The current dense wavelength division multiplexing (DWDM) based optical transport network architectures operate with fixed-grid employing fixed data rates. So, making this rigid approach to capacity allocation leads to inefficiencies in both spectrum allocation and energy usage. Flexible (or elastic) optical transport networks with flexible-grid were proposed to improve bandwidth provisioning efficiencies. Such networks support adaptive line rates and OFDM-based optical transmission, thus, this will lead to lesser network elements deployed and, consequently an improvement in energy efficiency. Similarly, wireless networks, whose data traffic is mostly derived from device-to-device (D2D) communication and heterogeneous 5G cellular networks (HETNETs) have since made tremendous strides to further enhance bandwidth by way of overlaying multiple types of low power small cells in a high-power macro cell. They afford more opportunities to explore the potential cognition and cooperation diversities to improve spectral efficiency. Thus in this work, we focus on both architectural design and operation of wireless and optical transport networks coupled with resource allocation. A model joint all photonic and wireless transport network architecture framework is proposed and analyzed. The architecture's performance in servicing high-capacity mobile back-haul and front-haul traffic and real-time services support is evaluated by both analytical and simulation approaches. Various routing and switching scenarios are considered. Overall, results demonstrate that elasticity allocation of resources (bandwidth) can vastly improve the network performance in terms of spectral efficiency, reduced locking probability, and enhanced end-to-end network throughput.

List of Figures

List of Tables

List of Abbreviations

BBU	Baseband Unit
BER	Bit Error Rate
BV	Bandwidth Variable
CAPEX	Capital Expenditure
CO-OFDM	Coherent Orthogonal Frequency Division Multiplexing
CB	Control Burst
CPRI	Common Public Radio Interface
CRAN	Cloud/Centralised Radio Access Network
D2D	Device to Device
DB	Data Burst
DEMUX	De-Multiplexer
DWDM	Dense Wavelength Division Multiplexing
EDFA	Erbium Doped Fiber Amplifier
EPC	Evolved Packet Core
EPON	Ethernet Passive Optical Network
GPON	Gigabit Passive Optical Network
HPN	High Power Node
ICT	Information Communication Technology
ITU	International Telecommunication Union
IoT	Internet of Things
IP	Internet Protocol
LTE	Long Term Evolution
M2M	Machine to Machine
MEC	Mobile Edge Computing
MLR	Mixed Line Rate
MIMO	Multiple Input Multiple Output
MUX	Multiplexer
NF	Noise Figure
NFV	Network Functions Virtualization
NRZ	Non Return to Zero
OBS	Optical Burst Switching
OEO	Optical-Electronic-Optical
OFDM	Orthogonal Frequency Division Multiplexing

OPEX	Operational Expenditure
OOK	On-Off Keying
OTN	Optical Transport Network
OXC	Optical Cross Connect
OSNR	Optical Signal to Noise Ratio
PON	Passive Optical Network
PSK	Phase Shift Keying
QAM	Quadrature Amplitude Modulation
QoE	Quality of Experience
QoS	Quality of Service
QoT	Quality of Transmission
QPSK	Quadrature Phase Shift Keying
RRH	Remote Radio Head
RAN	Radio Access Network
RMLSA	Routing, Modulation Level and Spectrum Allocation
SDN	Software Defined Networking
SLR	Single Line Rate
SMF	Single Mode Fiber
TDM	Time Division Multiplexing
TCP	Transport Control Protocol
WDM	Wavelength Division Multiplexing
WTN	Wireless Transport Network
XG-PON	Next Generation Passive Optical Network

1. Introduction

1.1. Background

The currently deployed 5G system is expected to completely change the way current networks operate in which data access is available anytime, anywhere, and to anyone or anything. 5G is designed to meet demands in terms of bandwidth, throughput, and latency, this is because of the tremendous increase of connected devices, smartphone users, wearables and most importantly IoT. With all these high bandwidth demands brought forward, providing high throughput, improved coverage, as well as ultra-low latency, is becoming more of an obstacle every day. From 1G to 4G-LTE, each upgrade from one generation of cellular network to another gave rise to power consumption, this is due to generations becoming more complex with additional hardware to cater for new demands or services and requirements. Moreover, since 5G network is designed to provide even higher data rates, it makes the 5G network even hungrier for energy and bandwidth. Energy consumption of a fully operating 5G wireless network is expected to be four times more than of 4G LTE, thus power efficiency is one of the prime factors in the design of current and future transport networks (wireless/optical) as compared to earlier generations where more focus was based on throughput and coverage. Figure 1.1 represents the predicted increase of power consumption in telecommunication networks and future trends.

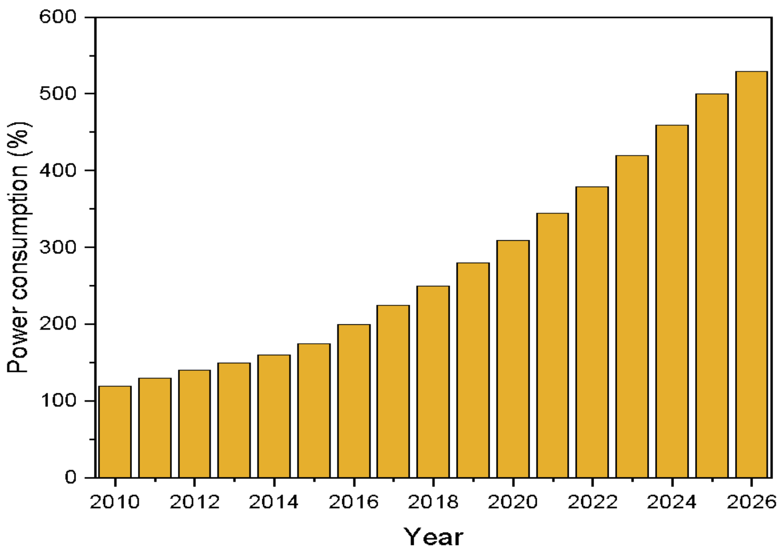


Figure 1. 1 Power consumption trend in telecommunication networks.

Various authors have proposed significant enabling technologies for 5G in the literature. These technologies aim to satisfy all the requirements for 5G networks and beyond. The technologies to be employed in the backbone network are Software Define Networking (SDN) and Network Functions Virtualization (NFV). Additionally, technologies to be employed in the access network are approaches such as heterogeneous network (HET-NETs) architectures, Ultra-Dense Network (UDN) architectures, millimeter wave (mmWave) transmissions, massive MIMO, Device to Device (D2D) communication Cloud/Centralized Radio Access Network (CRAN) and Mobile Edge Computing (MEC) are also under consideration. The deployment of these new enabling techniques creates several issues in terms of power efficiency, for example, a HETNET is densified with both low power base stations and high power base stations resulting in the overall increase of the total energy consumed by the network, The integration of SDN in the core network means more computational requirements by the SDN controller which may result in latency problems or even result in higher computational power in ultra-dense areas where more processing is required quickly and efficiently. Therefore an efficient transport network design is absolutely vital with the incorporation of all the enabling technologies. Energy efficiency must be achieved at all network levels. Finally, although most of the overall network energy is consumed by the RAN as shown in Figure 1.2, the operation of the core backbone network is also crucial for achieving energy efficiency. Therefore, energy efficiency architectures are required for both the backbone core network and the radio access network to achieve a holistic energy-efficient transport network. Figure 1.2 also shows energy consumption breakdown by various networking elements [2].

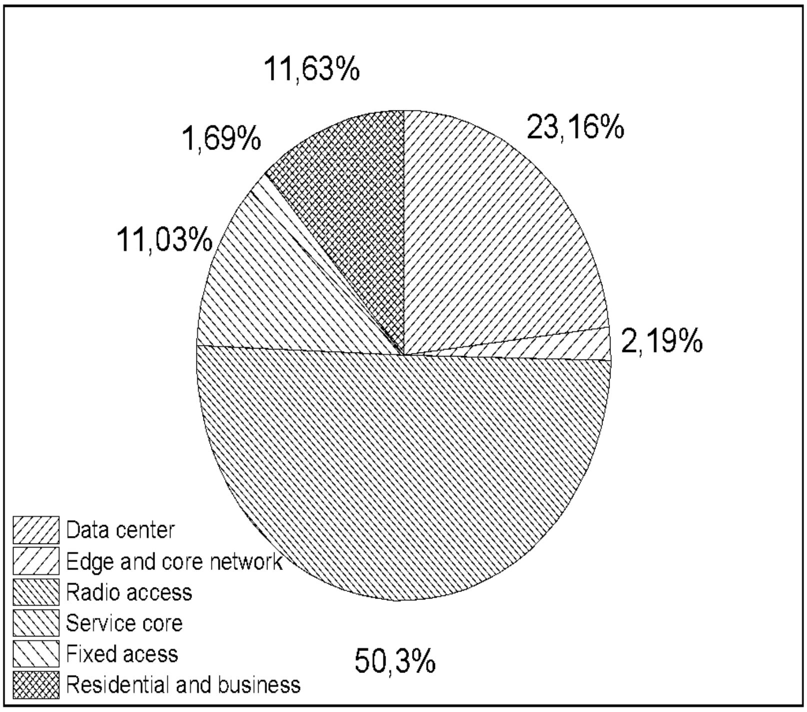


Figure 1. 2 Energy consumption breakdown by network elements

Furthermore, optical transport networks have proven to significantly reduce energy consumption, hence it is a necessity that the current and future wireless network must use optical transport networks in the mobile back-haul and front-haul in order to have a major recognizable reduction in electricity bill in the wireless communication industry. Efficient transport network is essential for every network. Traditionally, the concept of overprovisioning transport infrastructure has been adopted by mobile operators, this is due to advantages such as relatively reduced complexity because dynamic management of transport resources is unnecessary, but the issue with this approach is poor cost management in terms of equipment and further suffers from poor transport resource provisioning [3]. Hence the concept of flexible transport networks is absolutely pivotal to current networks such as 5G and future wireless or wired networks in general. A flexible transport network is advantageous because it is designed in such a way that transport resources are allocated on demand with varying traffic requests. This is possible through NFV, dynamic resource sharing and mixed line rates (MLR) [3]. Dynamic resource sharing basically means the sharing of transport resources for different purposes, and NFV can create a flexible platform by dynamically placing network functions in different locations based on specific demands of services. Furthermore, dynamic resource sharing with NFV can also mean migrating some of the complexities from the forwarding plane to the control plane, hence initiating SDN as part of the current and future wireless network designs.

Shortly, the requirement for an intelligent network that can take decisions intelligently improves network operation, energy efficiency, and scalability. With such a network, it can serve major benefits to economic concerns by reducing operating costs (OPEX). Finally, energy consumption and carbon footprint reduction have become essential key performance indicators (KPIs) in designing and evaluating wireless access networks. This is because most of the current wireless communication systems are energized by carbon-based power sources and currently the ICT systems are responsible for approximately 5% of the world's CO2 emission [4], and this number will continue to grow as the number of connected devices increase drastically. For that and other reasons, 5G and future wireless networks have made it one of its prime concerns to go green on its design, i.e., using green components and devices, green routing, renewable energy use, and efficient network design.

1.2. Research Problem

Previous generations from 1G to 4G have based their designs on improving throughput and extended coverage but not so much on energy efficiency and optimized resource allocation. The energy consumption of 5G networks is expected to be four times that of 4G [5]. Moreover, mobile networks alone consume about 0.5% of worldwide wide energy [6]. Ericsson Mobility Report predicts that by 2025, the aggregated end user data will increase four times compared to today's network [7]. Therefore, this negligence of power consumption of mobile communication networks cannot continue any further, the issue of energy efficiency needs to be addressed and resolved before any further damage shortly. By 2025, the ICT industry could be accountable for 30% of the total energy consumption globally, and data centers alone could be responsible for 3% CO_2 emissions [8]. Consequently, an immediate approach to reducing power consumption in transport networks is absolutely pivotal and requires immediate attention. Transport network architectures must be scalable enough to cater to various services and applications.

1.3. Thesis Objectives and Contributions

5G network is envisioned to consist of various diverse ingredients, termed heterogeneities [9], and merging them into a single holistic network [10]. Hence different frequency allocations for different cell coverage areas is considered, in which low frequency band aims to improve connectivity and mobility management, whereas higher frequency band aims to boost user data rates [11]. The heterogeneity introduced by 5G network results in two drawbacks. Firstly, it limits the use of 5G technology in a uniform way, basically preventing the broad use of 5G technology while increasing OPEX. Secondly, it makes the 5G network structure to be more complicated, and hard to monitor the entire network operation [12].

The work aims to ensure that 5G and future networks offer comparably improved elasticity, cost effectiveness, elasticity and spectral efficiency, programmability, optimum resource allocation, and energy efficiency. Furthermore, a greener design approach for wireless networks is investigated/ explored. To meet this goal, NFV separates network functions from dedicated proprietary hardware and SDN) which demarcates control and data planes in the network's interior to allow full programmability will be the pillar approaches. The ultimate objective is to improve the automation of the overall wireless/optical network management and control and flexibility in the optical domain. Lastly, an energy-efficient operation of the broad network is one of the main focuses, which includes intelligent adaptive topologies and efficient routing. Summarily the objectives are:

- Designing an enabling transport network architecture based on SDN as well as NFV approaches and paradigms.
- Performance analysis of the designed architecture by way of both analytical as well as simulation modelling.
- Designing and implementing of SDN and NFV based automation of the overall network management operations.
- Performance analysis on flexible optical transport networks for the currently deployed 5G networks and future networks.

1.4. Thesis Outline

The remaining sections of this thesis are arranged as follows.

In *Chapter 2*, a review of both optical and wireless transport networks is given. A descriptive overview of the main concepts, building components and network design trends for these network types is given.

In *Chapter 3*, we focus on the architectural design approaches for both optical and wireless transport networks that promote overall energy efficient operation. We will seek to qualitatively define an end-to-end power consumption model for both types of networks.

Chapter 4, focuses on the operational modes as well as resources allocation in terms of promoting energy efficiency for both optical and wireless transport networks. Specifically, we will look at an example of elastic (flexible) optical transport networks and equivalents (similar) wireless networks.

Chapter 5 focuses on energy efficient regenerator placement strategies on translucent optical networks.

In *Chapter 6,* we will further motivate the significance of energy efficiency in telecommunications and we will focus on the energy-efficient design and operation of both the optical and wireless transport network to achieve a proposed suitable architecture (joint all-photonic and wireless architecture) for the current 5G network and beyond. We will then look at OBS as the main switching paradigm for the proposed architecture and further analyze the overall network performance.

Finally, in *Chapter 7,* we finalize and conclude the research. We summarize the work presented in this thesis through contributions and potential applicability of the proposed architecture approaches, and finally point out directions for future research to extend on the presented work.

1.5. Summary Chapter Conclusions

In this chapter, we discussed the significance of energy efficiency and motivated by the need to design a transport network that will cope with future bandwidth demands (i.e. higher data rates, a higher number of end users, and increased energy consumption). The significance of designing an architecture based on a flexible optical transport network as well as wireless transport network is vital because as mentioned, flexibility allows for on-demand support of transport resources hence a decrease of wasted resources in the network, which results in the curbing of energy consumption and improved resources allocation. Finally, initiating new technologies such as SDN and NFV in the wireless/optical network will result in new strides towards achieving a holistic energy-efficient network.

2. Principles of Optical and Wireless Transport Networks.

2.1 Optical Transport Network (OTN)

OTN is a fixed access network using fiber optics ‘light’ which is already adopted by a variety of backbone networks due to its advantage of providing ultra-high capacity transport services. Networks such as 5G and beyond will require an upgrade from conventional back-haul/front-haul to cater for 5G mobile services. Therefore, a study on OTN is essential in order to provide the necessary upgrade from current network architectures to future networks architectures.

2.1.1 Introduction

As the ever increasing internet traffic demand due to bandwidth extensive and bandwidth dynamic applications (e.g. IoT, online gaming, YouTube and cloud computing etc.), operators of telecommunications are deeply researching on exploring new design approaches which are alternatives to improve and upgrade on their current transport network in order to cope with the ever increasing internet traffic demand. Herein the term “transport” has been adopted in this work in a sense of representing improved high-performance links connected across geographical regions, even over long and isolated areas [13]. With the surging of internet traffic, so does the energy consumption of transport networks, thus more attention is focused on energy efficiency as a design factor for current and future network design and operation. Moreover, a direct upgrade on a network capacity usually coincide with an increase in power consumption which then results not only increasing OPEX but also plays a major role towards carbon emission. Therefore, towards a process of curbing this effect, an efficient optical as well as efficient wireless transport network will play an essential role to cater for new capacity requirements and also minimize energy consumptions. Additionally, Commercial transport networks of 10 Gbps has been upgraded and increased to 40 Gbps and/or 100 Gbps. However, practically this upgrade of data rates capacity from 10 Gbps may not necessarily mean that each and every service now needs a transmission rate of 40 Gbps and/or 100 Gbps, and in such a situation, this upgrade from 10 Gbps to 40 Gbps and/or 100 Gbps will result in poor utilization of network resources [14]. Traditional optical networks are based on

Wavelength Division Multiplexing (WDM) technologies, which generally operate at single rates (SLR), however, mixed line rate (MLR) was introduced to improve network flexibility to support various traffic requests with different rate requirements [14]. Finally, for a long-term solution an innovative flexible/elastic-grid network approach based on Orthogonal Frequency Division Multiplexing (OFDM) promises to be an interesting solution, since it permits for the adaptation of the channel bandwidth according to the requested demand, as compared to the conventional fixed-grid Wavelength Division Multiplexing (WDM) with SLR, this will be described in detail later.

2.1.2 Wavelength Division Multiplexing

Wavelength Division Multiplexing (WDM) is an approach used in the optical transport network to utilize the bandwidth capacity of a transmission channel by simultaneously transmitting optical signals on the same fiber, exploiting wavelength multiplexing. Each optical signal is mapped to a separate wavelength channel. The various channels are then multiplexed at the transmitter side and ultimately de-multiplexed at the receiver side. This approach is illustrated in Figure 2.1.

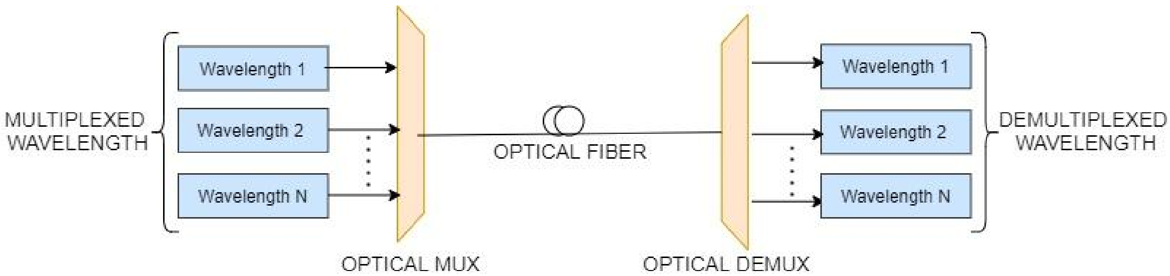


Figure 2. 1 General WDM system

Telecommunication operators have adopted this technique as the standard transport medium on core (backbone) networks, which resulted in the standardization of WDM systems. WDM systems were standardized by the International Telecommunication Union (ITU), so that different telecommunication vendors can inter-operate their equipment easily on the network backbone. The ITU standardized and differentiated each band according to their wavelength range as depicted in table 2.1.

Table 2. 1 ITU standardized wavelength range

Band Name	Wavelength Range
O-band	1260-1360 nm
E-band	1360-1460 nm
S-band	1460-1530 nm
C-band	1530-1560 nm
L-band	1560-1630 nm

However, the widely used frequency band in long distance optical transmission is the C-band due to its advantage of low attenuation per kilometer i.e., typically 0.2 dB/km, and also the possibility of using optical amplifiers (EDFA) is another major advantage of the C-band for long distance transmission. In contrast with the other bands which offer slightly higher attenuation loss per kilometer and amplifiers cannot be used [15].

Dense Wavelength Division Multiplexing (DWDM) refers to a denser packing of channels in the C-band to get the lowest signal attenuation in this band and also leverages on using Erbium Doped Fiber Amplifiers (EDFA). This standardization of transmitting in the C-band in which each channel is allocated a fixed 50 GHz frequency slot is referred to as the ITU wavelength grid which is illustrated by Figure 2.2.

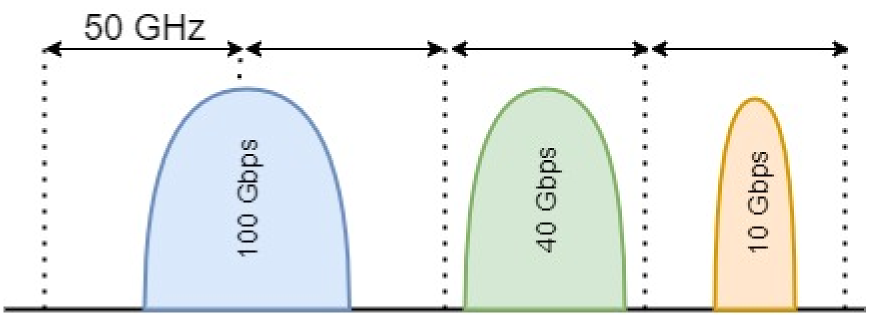


Figure 2. 2 ITU wavelength grid.

Conventional DWDM networks operate with the ITU grid of 50 GHz frequency slots which support channel capacity of up to 100 Gbps. As mentioned before these wavelengths (channels) are multiplexed and the composite signal is transmitted over the optical fiber, along the path at intermediate nodes, the channels can either be passed through the node or regenerated depending on the signal strength. The intermediate nodes along the path can be referred to as optical cross connects (OXCs). Lastly, optical amplifiers can also be used along the fiber path to boost the signal for longer distance transmissions.

2.1.3 Single and Mixed Line Rates on WDM Networks

Legacy WDM networks are designed based on 10G channel rates, which is also known as Single Line Rate (SLR), however this single line rate is outdated and inefficient as internet traffic demands are becoming more heterogeneous to support higher channel rates such as 40G and 100G on the same link. As the name describes, SLR operates at a single capacity despite the requirements or demands of the network, for example a single line rate of 100G can be used despite different capacity demands which actually exists between nodes whether less or greater than 100G. Basically in a case where the network traffic demand between nodes is 10G, if a fixed line rate of 100G was deployed then that will result in poor utilization of resources, hence energy will be used inefficiently to support a 100G line whereas a 10G line could have been used, noting that a higher-rate transmission basically consumes higher power compared to a lower-rate transmission. This inefficiency of line rates in legacy WDM networks has motivated for a more flexible and adapting technology to support not only a single line rate but different line rates depending on the amount of traffic, this technique is referred to as mixed line rate (MLR). Furthermore, this technique offers higher resource allocation which results in a far more energy efficient transport network. MLR have the ability to adjust and adapt the network bandwidth to a specific network demand, hence MLR is also referred to as elastic optical network.

2.1.4 Building Components for Optical Transport Network (OTN)

Successful optical transmission of data is dependent on few basic components or block along the optical path, which play significant role in the transmission and reception of optical signals over long-haul networks. Namely, these blocks are optical transponders, optical fiber, optical amplifiers and optical cross connects (OXC) as represented in Figure 2.3.

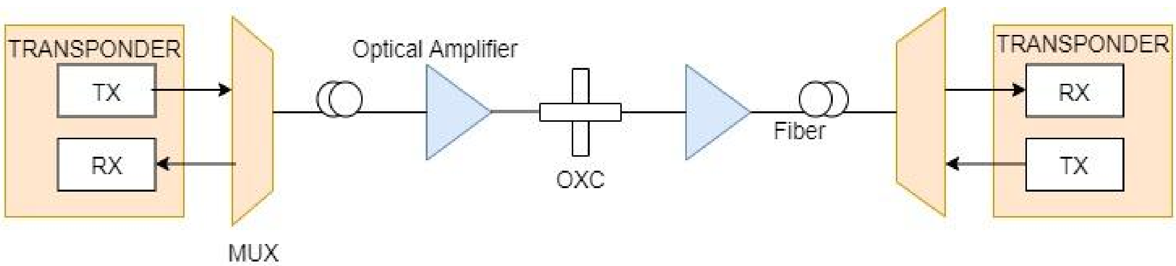


Figure 2. 3 Building blocks for optical transport network

2.1.4.1 Optical Transponders

A transponder performs a significant role in the transmission and reception of optical signals. The transponder is responsible for generating the optical signal with a specific wavelength under the C-band. Moreover, transponders also perform functions such as optical-electrical conversions, electrical-optical conversions and applying modulation/demodulation techniques (OOK, QPSK, QAM) which suit the desired transmission reach and line rates. WDM transponders assign each signal a fixed channel/frequency slot of 50 GHz as per standardized in the ITU-T grid. Transponders have the ability to detect (demodulate) the optical signal using a single photo diode at the receiver side, this method is known as direct detection and mostly used for transmission speed up to 40 Gbps. The other form of demodulation performed by transponders is coherent detection which is used for dual polarization schemes such as dual polarization QPSK. All the mentioned tasks performed by transponders make them high consumers of energy in the optical transport network. Figure 2.4 shows the generic block diagram of a transponder.

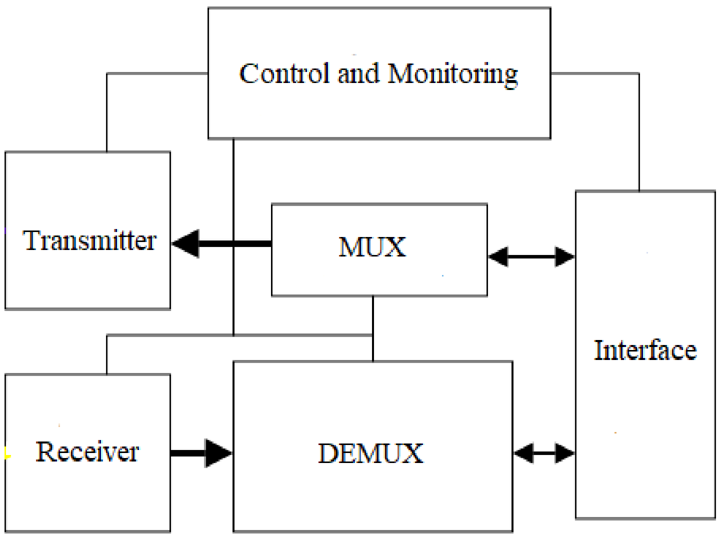


Figure 2. 4 Typical block diagram of a transponder.

2.1.4.2 Optical Fiber

Optical fibers in their simplest form can be viewed as structured in a coaxial arrangement of two homogeneous glasses (core and cladding) as depicted in Figure 2.5. In a more detailed definition, optical fibers consist of inhomogeneous core surrounded by a cladding region and shielded by a plastic jacket [17]. Optical fibers can support a number of guided waveforms called modes i.e. single mode fiber (SMF) are significantly used in communication where high bandwidth is required over long distances. A single mode fiber only support one waveform, while multimode fiber (MMF) support over hundreds of modes (waveforms). Optical fibers have the ability to act as waveguide, and transmit or guide light waves from source to destination using a principle of total internal reflections at the core-cladding interface. Optical fiber suffers some physical impairments known as attenuation and dispersion, which eventually destroys the optical signal to an un-recoverable signal strength over long distances if not amplified along the path. Attenuation in a fiber path is a loss of wavelength power over a specific distance, and dispersion results from optical pulses (light pulses) being wider over long distances, which than affects neighboring signals and causing inter-symbol interference. Due to these impairments, optical amplifiers are used at intermediate locations (nodes) [16]. Despite the physical impairments, optical fibers are still more advantageous over traditional copper since it offers much broader bandwidth and is much less prone to electromagnetic energy interference than copper.

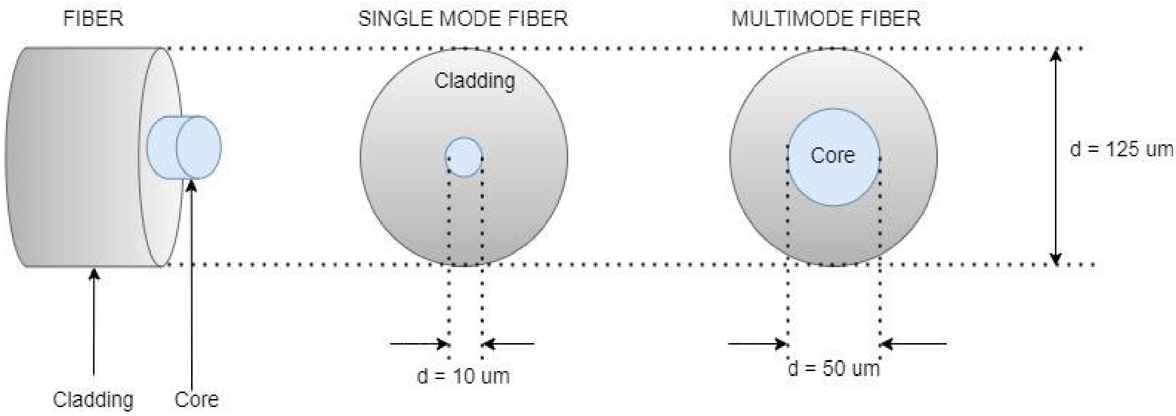


Figure 2. 5 SMF and MMF

2.1.4.3 Optical Amplifiers

Optical amplifiers play a significant role in long distance transmission, in contract with regenerators, optical amplifiers have the ability to improve signal power levels without any optical-electrical conversion taking place, which is a huge advantage on less processing time and decreased power consumption. These amplifiers are used to extend transmission reach of optical signal and can amplify every signal contained in the C-band. The most widely used amplifier is erbium doped fiber amplifier (EDFA), which is made up of optical fiber doped with an earth element erbium (erbium doped optical fiber) as shown in Figure 2.6. These EFDA amplifiers offer a reasonable noise figure (NF) ranging between 4-6 dB. The other type of amplifier is roman amplifier, which is not widely used because of higher cost despite having much less NF as compared to EDFA.

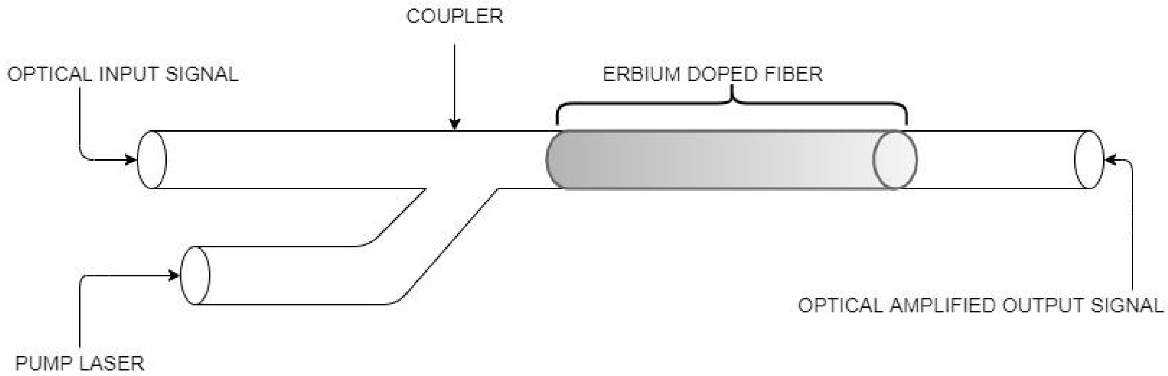


Figure 2. 6 Erbium doped fiber amplifier

2.1.4.4 Optical Cross Connect

Optical cross connects (OXCs) are nodes connected to more than one node, these are responsible for switching of optical signals at a different wavelength from an input port to a particular output port. OXCs can switch or direct optical signals (light waves) between different ports directly on their optical form, as shown in Figure 2.7. OXCs are flexible compared to legacy SDH switching mechanisms since signal switching is done at optical level. Furthermore, OXCs can switch (add/drop) channels at frequencies specified by the ITU-T grid which is 50 GHz. Future OXCs are expected to have a variable bandwidth to accommodate future elastic transport networks.

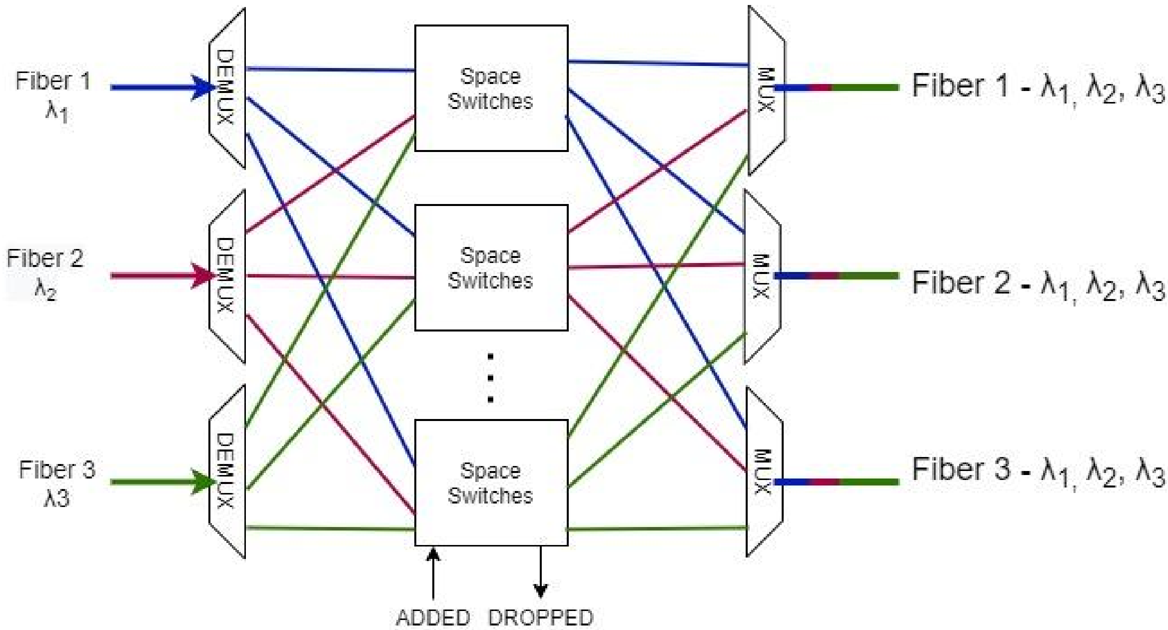


Figure 2. 7 Block diagram of optical cross-connect

2.1.5 Transport Network Regeneration Allocation

Optical transport networks can be divided into three variants depending on the allocations or presence of intermediate regenerator nodes in the light paths. These variants are transparent, translucent and opaque architectures.

2.1.5.1 Transparent Optical Network

In a transparent optical network, the nodes which are immediately located between two communicating sources can be avoided and bypassed, hence less conversion takes place resulting in less signal processing and less power consumption because regenerators are bypassed. The only drawback with this approach is that geographical reach (optical reach) is limited as maximum optical reach of an optical signal is limited due to signal degradation and impairments if no amplification/regeneration takes place.

2.1.5.2 Translucent Optical Network

Translucent optical network aims to overcome the issue of geographical reach occurring on a transparent network, this is achieved by deploying signal regenerators at different nodes but only where regenerators are necessarily required to extend optical reach. This means an optical signal travelling from source to destination can be regenerated several times at intermediate nodes depending on whether regeneration is required or not. The number of regenerators on a light path depend on the optical reach and the length of the light path.

2.1.5.3 Opaque Optical Network

Opaque optical networks on the other hand can be viewed as an inefficient approach because from source to destination there will be conversion interfaces at all nodes, hence regeneration occurs at every node even if the signal does not require any amplification/regeneration, this essentially means power is not utilized properly in this technique. It has been proven that the main contributors towards the overall energy consumption of the entire network is due to optical-to-electrical (O-E) conversions and/or electrical-to-optical (E-O) conversions which is done at the regenerator nodes. This means since opaque optical networks have conversions and regenerators at every node then the total energy consumption of the overall opaque network will increase significantly.

2.1.6 Passive Optical Networks (PONs)

As the name implies, PON technology can be described as a telecommunication technology which utilizes passive components such as combiners, splitters, and couplers to form an optical point-to-multipoint architecture. This architecture is bidirectional (upstream/downstream), and it offers key advantages such as improved space utilization, transparency, security and less energy consumption. Energy is consumed less in this architecture based on that the PON technologies do not use active components such as amplifiers, thus PON does not require any electrical power resulting in an energy efficient operation [18]. A typical PON architecture is depicted in Figure 2.8. This architecture is made up of an optical line terminal (OLT) also referred to as a central office, a passive splitter also referred to as an optical distribution network (ODN), and several optical network units (ONUs). An OLT is situated in the central office and it is the endpoint of a PON but it is responsible for an interface to external networks such as core networks. An ODN functions as a link between the OLT and various ONUs, it comprises of passive splitters or WDM couplers or combiners which are responsible for distributing the optical signal from OLT to various ONUs. An ONU is responsible for converting the optical signal back to its electrical form and de-multiplexing it. There are different types of PON, namely, EPON, GPON, XG-PON, and NG-PON 2 as per IEEE standardization [22].

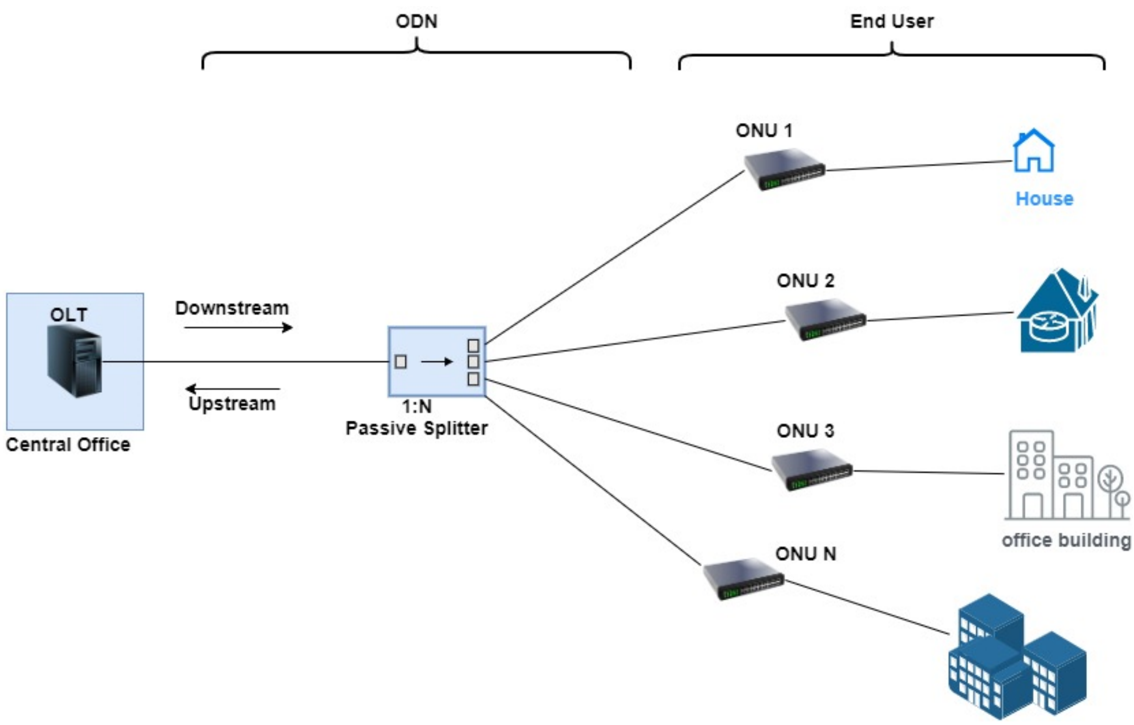


Figure 2. 8 PON architecture

Ethernet Passive Optical Network (EPON) is a distribution topology from one point to multiple points which transmit data from the central office to several ONUs (typically 16 ONUs per OLT) over a distance of 10 to 20 km through a passive optical splitter. EPON support transmission rate of 1 Gbps upstream and downstream. Gigabit Passive Optical Network (GPON) is a slight improvement from traditional EPON, it can support transmission rates of 2.5 Gbps downstream and 1.2 Gbps upstream over a distance of 60 km. Next generation Passive Optical Network (XG-PON) is a much more improved version of PON because it supports transmission rates of 10 Gbps for both downstream and upstream. However, Next Generation Passive Optical Network stage 2 (NG-PON 2) is a more advanced technology and a successor standard of EPON, GPON and XG-PON. It supports transmission rate of 40 Gbps downstream and 10 Gbps upstream, it also supports multiple wavelengths per direction and compatible over ODNs, and also it supports multiple multiplexing techniques such as time division and wavelength division multiplexing [18].

A. TDM-based PON

Time Division Multiplexing (TDM) is the traditionally used technique for multiplexing on a PON, where data is transmitted in packets and allocated a time delay to differentiate the packets directed to each ONU. On a TDM-based PON point to multipoint network architecture, the OLT transmit time-delayed packets to the ONUs through a splitter and optical fiber lines. This is depicted in Figure 2.9, where t1-t4 represent time delays.

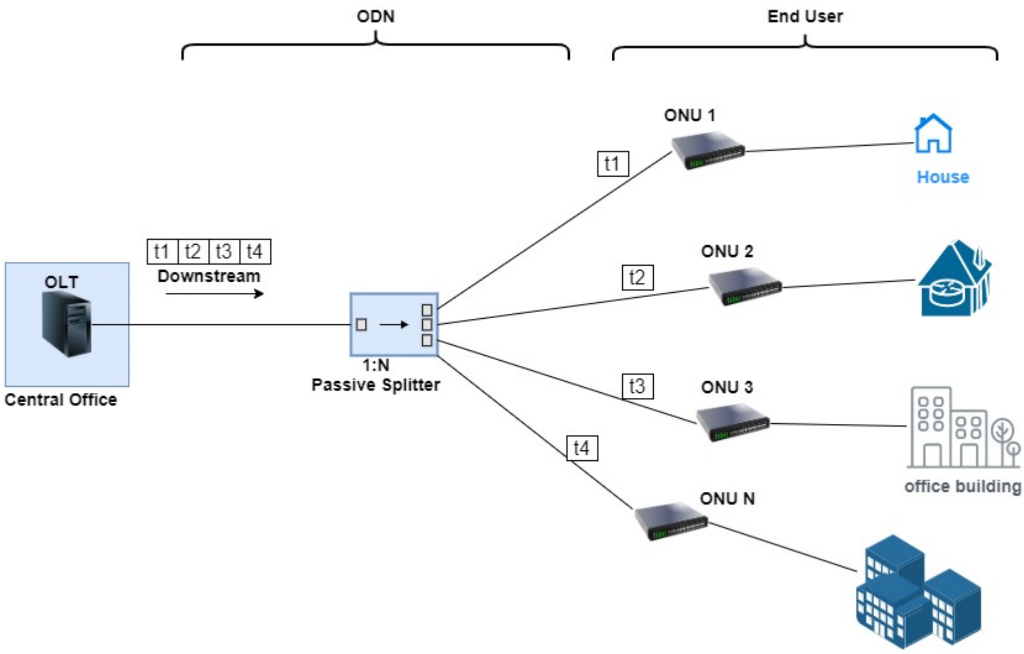


Figure 2. 9 TDM-based PON architecture

B. WDM-based PON

TDM-based PON is rather a non-optimum network based on the fact that it requires time synchronization and also it has a restricted bandwidth utilization. On the contrary to WDM-based PON, which allocates different wavelength to each end user which results in low transmission loss, improved mobility of the user and higher bandwidth [18]. In this approach wavelength-assigned data is transmitted from the central office to distributed ONUs through optical filters or WDM couplers. WDM-based PON is depicted in Figure 2.10.

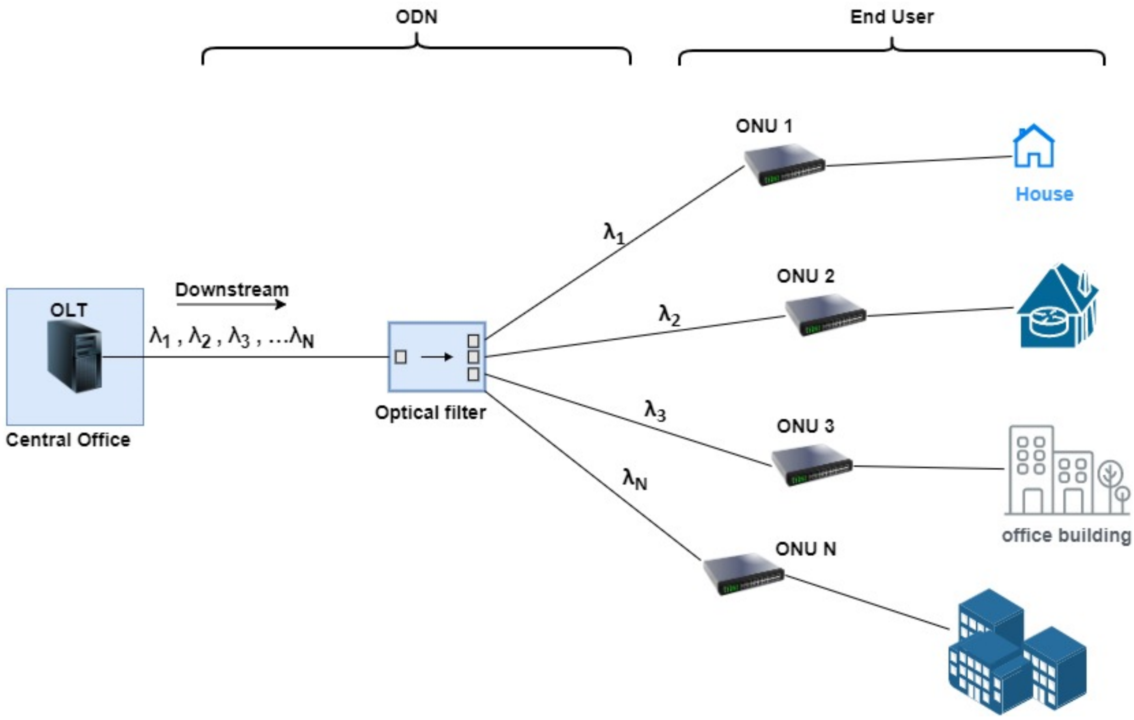


Figure 2. 10 WDM-based PON architecture

2.1.7 Quality-of-Service (QoS)

Survivability of a network is vital for any transport network design because it improves the QoS. Survivability refers to the network's ability to provide uninterrupted and improved services at an acceptable level despite the occurrence of different failure scenarios [19]. On the other hand, QoS is basically the network's ability to measure the overall performance of the network and provide optimum services to the network traffic over different technologies. QoS can also be defined as the ability of the network to enforce preferential treatment through classifications [19]. Optical networks or networks in general use QoS parameters to determine the overall performance of the network. The most commonly used QoS parameters are Bit Error Rate (BER), packet loss, survivability and reliability, response times and delays, and fault-tolerance. Optical fibers are less prone to interference as compared to co-axial cable or copper wire, but still fiber optics suffer from certain impairments such as dispersion which then results in an unacceptable QoS. To determine a signal quality, a measurement of BER is used in networking. BER can be expressed as:

$$BER = \frac{Number\ of\ error\ bits\ recieved}{Total\ number\ of\ transmitted\ bits} \tag{2.1}$$

BER measurement is significant because in a network many errors may occur resulting in packet of data getting lost along the path, or maybe the signal degrades to an unrecoverable signal, or even there might also be a fault in the hardware path. All the aforementioned errors may occur any time in a network which then affects the QoS. Finally, issues such as delays can also affect the QoS. Delays are regarded as the time it takes (in milliseconds) for a bit of data to be moved from one point to another. Delays can be divided as follows [20]:

- Propagation delay – time duration for a data packet to travel through a fiber optic or wire.
- Processing delay – duration of time in which a data packet is processed in network system.
- Transmission delay – duration of time that is needed to send all bits of data into a medium (fiber optics/wire).

Finally, the reliability and survivability of the optical network impacts the QoS in a sense that the network should be resilient to overcome unexpected failures and have the ability to withstand unexpected malfunctioning. Traffic engineering refers to an engineering of network resources to achieve optimum performance in terms of network operation, while also enhancing the reliability of the network against any malfunction. The concept of traffic engineering (TE) is essential in a network to fully optimize the network performance and maximize the throughput.

2.2 Wireless Transport Network (WTN)

Wireless technology can be essentially defined as an approach of transferring data between two or more devices without the use of physical wires, and this process can be achieved through using radio and satellite communication. Radio is fundamentally the wireless transmission and reception of electric waves by means of electromagnetic waves [21]. Wireless access networks have developed in recent years to a point where wireless networks can provide comparatively similar services as fixed access networks, however, the difference is that the transmission in WTN is done wirelessly through the air. Therefore, wireless transport network can be viewed as wirelessly connected links across a geographical area sharing information, whether transmitting or receiving data over long or short distances as shown in Figure 2.11. This communication or sharing of information wirelessly is fundamentally achieved by means of Mobile IP core, base stations and cells. The core network can be viewed as a gateway to connect to other networks, and also responsible for providing management entities to manage the wireless communication.

2.2.1 Background Information

The Third Generation Partnership Project (3GPP) and other standardizing bodies have adopted some standards of the wireless access technology in-order to achieve optimal operation of the wireless access network. A basic architecture of a wireless access network comprises of wireless end users that are wirelessly connected to the base stations, and those base stations are further connected to backbone network via a backhaul network. This architecture of a wireless access network is typically illustrated in Figure 2.11.

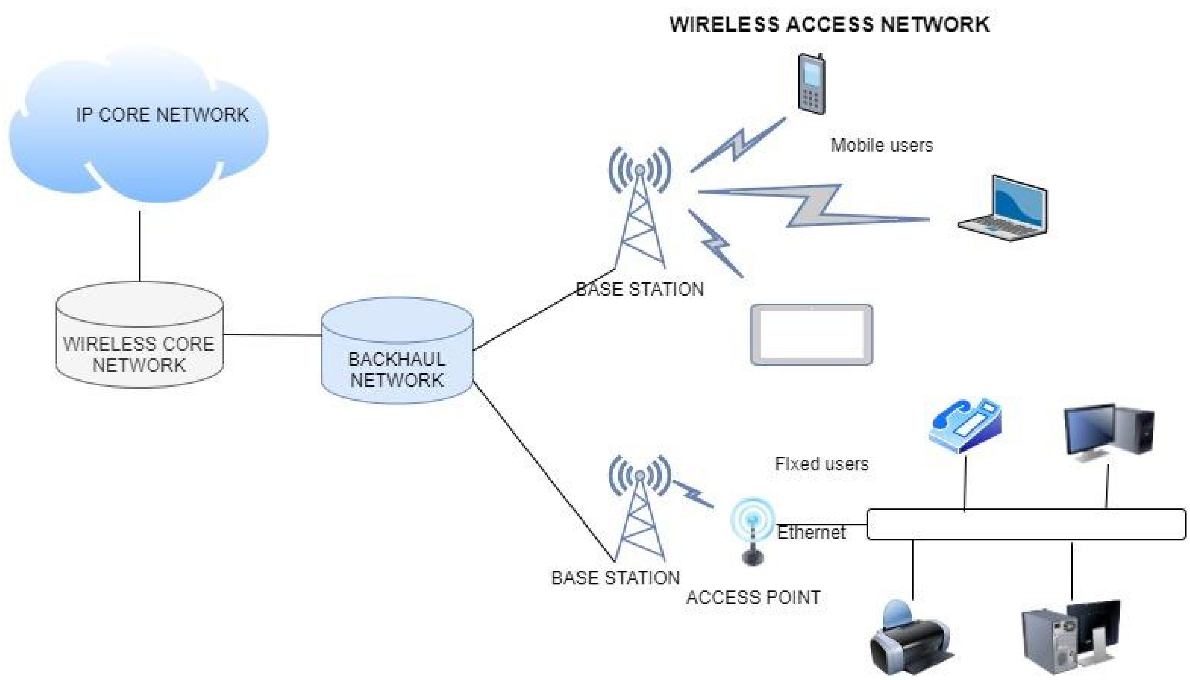
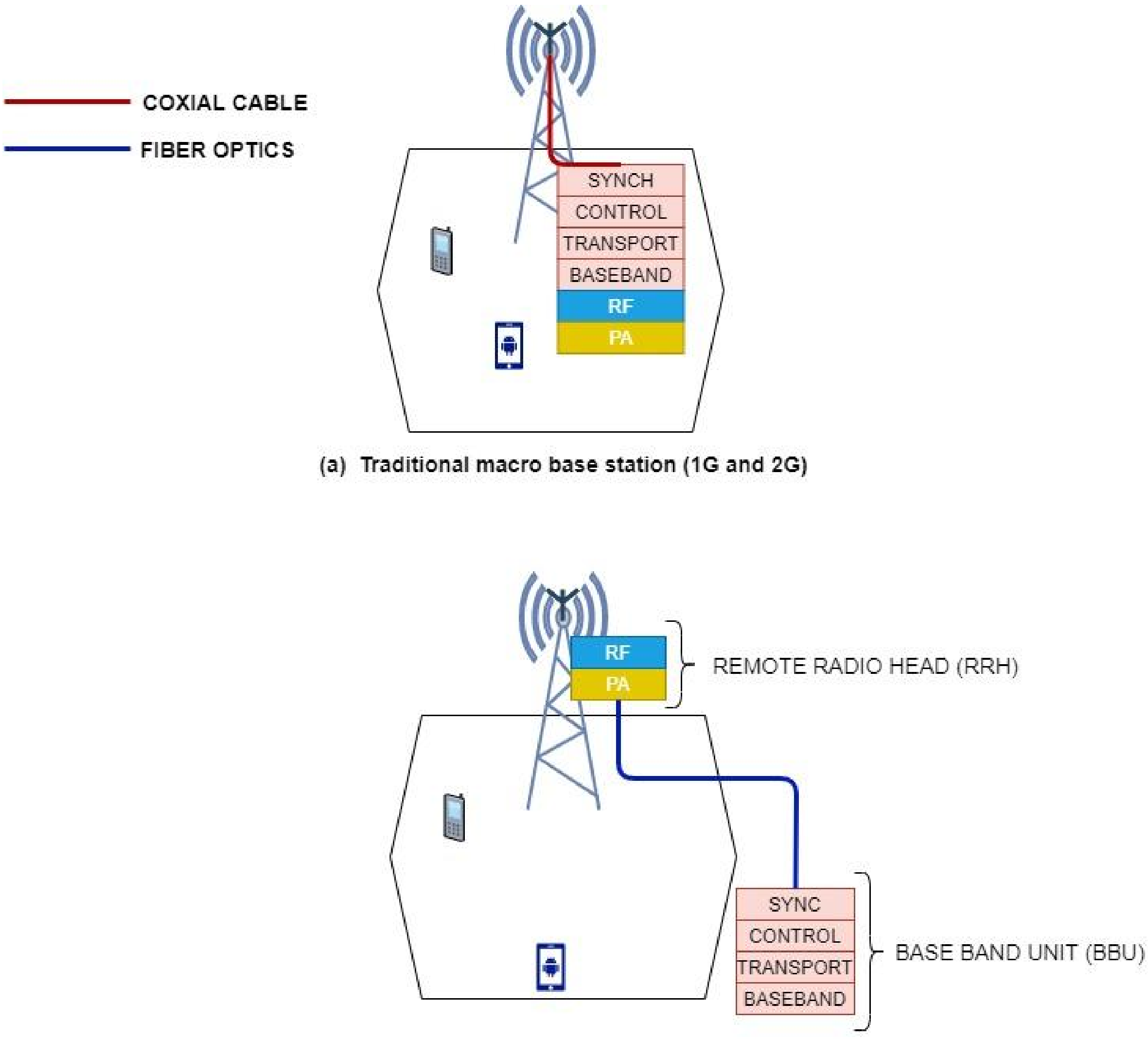


Figure 2. 11 Wireless access network architecture [22]

Wireless networks are commonly referred to as mobile networks or cellular networks. Cellular networks in a sense that an area in which a network base station covers is split into cells. Generally, in a mobile network, users can communicate with each other via a base station that serves coverage of that specific area. The base station is primarily designed for baseband processing i.e. coding, modulation, Fast Fourier Transform (FFT) and radio functionalities such as frequency filtering, digital processing and power amplification.

The cellular network architecture comprises of two parts which is the Radio Access Network (RAN) and the backbone core network. Cellular networks operate based on three wireless broadband access technologies, namely Long-term evolution (LTE), Wi-Fi and mobile WiMAX. Firstly, looking at LTE networks which are based on an all IP network, and LTE uses a mobile core termed the evolved packet core (EPC) that is made up of the serving gateway (S-GW) and the packet data gateway (P-GW). Additionally, LTE comprises of a RAN that is made up of special base station known as evolved-node base stations (eNodeBs). Secondly, Wi-Fi is basically a wireless local area network, which can reach a smaller coverage as compared to LTE, however, it is highly efficient in data rates hence higher bandwidth is allocated to end users [22]. Lastly, WiMAX is much similar to LTE technologies and it is based on standards set by the IEEE 802.16. Mobile WiMAX can support point-to-multi point modes as well as mesh modes as part of its architecture.

Traditional architectures such as 1G and 2G networks integrated baseband processing and RF functionalities within a single base station whereby the receiving antenna is also located close to the radio modules. Moreover, these traditional architectures used co-axial cables for connecting in-between modules, noting that co-axial cables exhibit high attenuation losses. This architecture is represented in Figure 2.12(a). However, 3G and 4G mobile networks opted to use a much-improved architecture whereby the base station is physically divided into two parts, the signal processing unit and the radio unit. The radio unit is generally called the Remote Radio Unit (RRU) or the Remote Radio Head (RRH), while the signal processing unit is referred to the Baseband Unit (BBU). Furthermore, this architecture used fiber optics between RRH and BBU modules to overcome high attenuation losses introduces by co-axial cables. This architecture is represented by Figure 2.12(b).



(a) Traditional macro base station (1G and 2G)

(b) Base station with RRH (3G and 4G)

Figure 2. 12 Evolution of base station architectures [23]

Today the world is currently deploying 5G wireless network, because of the increase in traffic on online devices, hence frequency band are becoming congested. This has prompted for new strategies that use new frequency range or frequency spectrum such as millimeter wave (mmWave). MmWave is an under-utilized spectrum between 30 GHz and 300 GHz, the issue with this frequency range is path loss, but it offers new advantages such as increased data rates which is an essential for 5G networks. The use of mmWave as part of 5G mobile network has introduced a dense deployment of small cells using low power access points. Ultra dense network is used to fulfil the need of highly congested areas which require more cell deployment. Conventional wireless transport network comprises of macro cells which support large area coverage using high power base stations, in contrast to ultra-densification of small cells which aims to increase the capacity of the network through the densification of low-power and low cost access points. This deployment of small cells on top of conventional macro-cells forms a heterogeneous 5G network. Small cells are divided into three types, i.e. pico cells which are used at a coverage range of 100m to increase capacity whether indoors or outdoors, femto cells which offer a much smaller range than pico cells of 10 m to 30 m coverage to also increase capacity, and lastly there are RRH which can only be deployed outdoors. RRH are typically smaller power base stations densely deployed to increase capacity. RRH are used in 4G networks, but the issue with RRH is interference and power consumption, intelligent techniques are described in chapter 3 on how to efficiently utilize energy in the densification deployment of cells. One way of doing so, is through the approach of centralized/cloud radio access network (C-RAN) which aims at reducing energy consumption and also reduce cost while optimizing a better management of the mobile network. This concept basically centralizes the base band unit/processing units in central office rather than traditional RAN where the baseband (BBU) are collocated at each and every base station site.

2.3 Summary Chapter Conclusions

In this chapter we discussed the fundamental concept of OTNs as well as WTNs. OTN is based on a fixed access network using conventional WDM strategies which transmit multiplexed wavelengths through the fiber medium. WDM is considered inefficient in today's network because of the use of a fixed ITU grid, thus new technologies such as MLR are implemented to improve elasticity in OTNs. In contrast, WTNs on the other hand are based on a wireless transmission using base stations and access points. WTNs give rise to new technologies such as C-RAN, which is a promising technology to cope with the recently deployed heterogeneous 5G network.

3. Energy Efficient Architectural Design Approaches in WTNs.

3.1 Introduction

Infrastructures and devices of the internet's core backbone network such as switches, routers, transmission systems etc. are estimated to consume approximately over 12 % of the entire internet energy usage [24]. Even though most of the power is consumed by the access network, also a huge amount of power is consumed by the core network which cannot be neglected, which is why it is also a necessity to design an efficient core network for 5G network and beyond. WTNs consist of a radio access network as well as a core backbone network. In this chapter energy efficiency is firstly achieved in the core network architecture then down to the radio access network architecture. 5G network and future networks will be the main focus. 5G networks are required to be more agile, resilient and autonomous. Moreover, 5G is required to operate in a green way whereby energy efficiency becomes the primary goal. In today's networks, backbone core network is based on Software Defined Networking and Network Functions Virtualization as part of its architectural infrastructure. To realize modern networks and implement an energy efficient core network, key techniques and approaches should be adopted on the currently deployed 5G networks and also future networks, namely, energy efficient re-design of the network which includes smart topology designs, as well as energy efficient operation which includes load adaptive operations and green routing. Green devices and components are essential in a network such as 5G to operate at the backbone core, these low power devices and components not only reduce power consumption but also they have low carbon emissions, which is why it is vital that copper based access and core network technologies are replaced with optical equivalents. In a Software Defined Networking (SDN) network architecture new possibilities emerge, such as a centralized intelligent control of the network using software applications whether manually or automatically. This approach can be applied to data centers and other network equipment in the core network. Furthermore, hardware components/devices are also consumers of energy, which means hardware solutions are of critical importance in energy efficiency at the network, hence the introduction of cloud-based implementation as well as virtualization to minimize hardware devices. Network Function Virtualization (NFV) mainly focuses on virtualizing network services in a technique to minimize the deployment of hardware in return minimizing power consumption of the overall network. With NFV approximately 30% of power consumption can be reduced with its implementation into future networks [25]. Finally, over and above the reduction of energy, NFV also has the potential to minimize OPEX by reducing the conventional purpose hardware upgrades [26], Since upgrades can be done via software.

3.2 Energy Efficient design approaches for backbone core networks.

The current 4G LTE architecture is consuming extremely high energy and if neglected it could go even higher as more devices are coming online, so this has prompted for a careful consideration and closer look at the architecture to reduce the overall power consumption and improve on resources allocation. The main functional entities of the 4G LTE architecture are the Evolved Packet Core (EPC), and the Evolved Universal Terrestrial Radio Access Network (E-UTRAN). The E-UTRAN is basically the wireless radio access network which basically comprise of nodes (i.e., base stations and antennas). In this section, the focus is on the backbone network on ways to improve network operation while minimizing power usage to achieve a fully energy efficient core network for current and future networks. The EPC is the core network which comprise of four main components, the Mobility Management Entity (MME), the Serving Gateway (S-GW), the Packet Data Network Gateway (P-GW), and the Home Subscriber Server (HSS). Basically the MME is responsible for mobility such as handoffs as well as network setup procedures, the S-GW is responsible for serving nodes and connecting them to the P-GW, the P-GW is responsible for creating a path to reach the Packet Data Network (PDN) or the internet, and finally the HSS represents the user database which serve information to the MME for user authentication. In order to improve the core network of 4G LTE to a greener and efficient backbone network to cope with future network services, a reduction in core components is essential as well providing a platform where virtualization is possible in the core, hence the introduction of SDN and NFV in the 5G backbone network and future networks is imperative. Generally, a core network can be regarded as the network that connects the radio access network to the internet as shown in Figure 3.1.

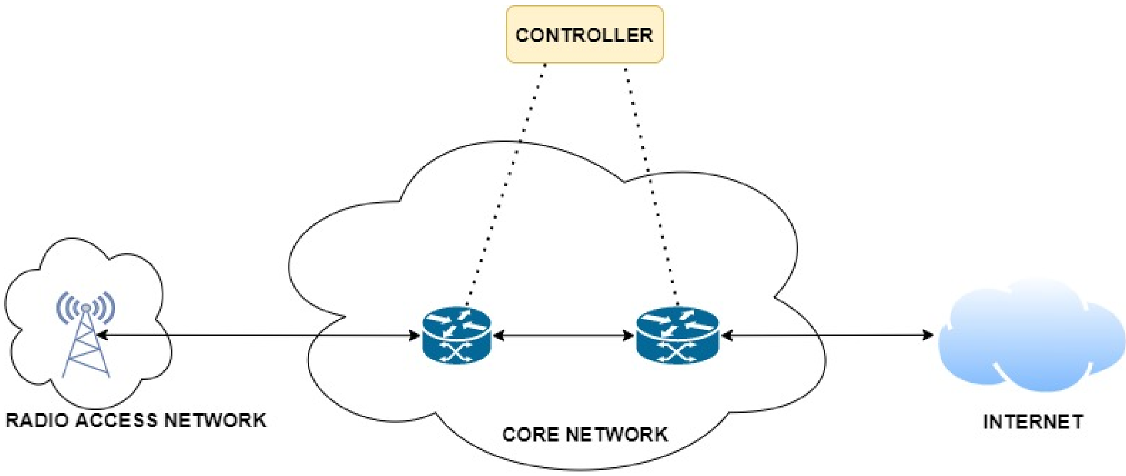


Figure 3. 1 General Core backbone network architecture

3.2.1 Software Defined Networking (SDN)

Many of the Information Communication Technology (ICT) Companies such as Cisco and Google have adopted software defined networking in their network infrastructure which includes data centers [27]. This is because SDN demarcates the general vertical integration by separating the control plane from the forwarding plane, thus improving flexibility as well as user experience. The demarcation of the two planes (data and control plane) opens new possibilities for current and future network design of architectures, because the control plane is regarded as the intelligence of the entire network while the data plane is merely regarded as forwarding devices. So with the intelligence stripped off from the forwarding devices (switches) because of this demarcation, switches begin to operate as pure forwarding devices with no intelligence as compared to traditional switching devices which were responsible for both forwarding and processing as shown in Figure 3.2. However, in SDN, a logically centralized controller is responsible for all the network intelligence which mainly includes controlling the traffic and processing. This method overcomes the traditional forwarding table format used on standard switches.

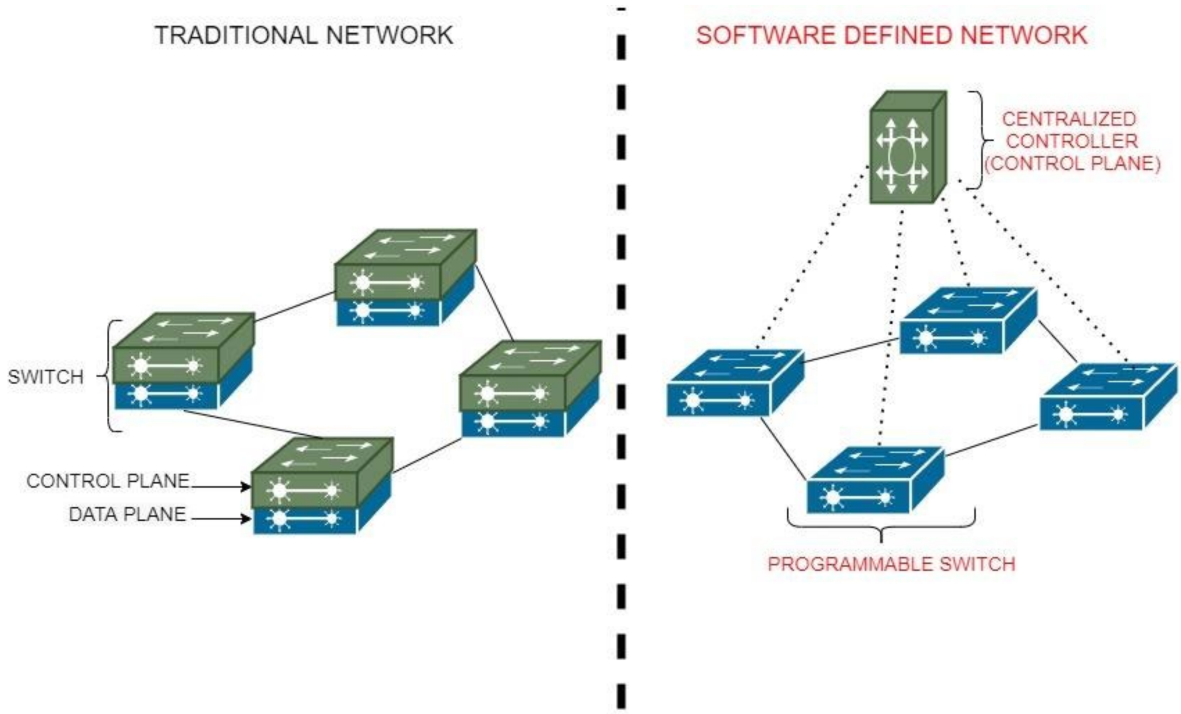


Figure 3. 2 Traditional networks vs software defined networks [28]

A more detailed architecture of SDN is presented in Figure 3.3. The controller termed SDN controller is connected with all switches termed SDN switches via a well-defined programmed interface termed Application Programming Interface (API). The widely used API is OpenFlow, which is regulated by ONF (open network foundation) [27]. OpenFlow can also be adopted in the core of networks such as 5G and beyond as the common medium of interface. There are quite a few advantages of SDN such as intelligent networking, implantation of virtualization of resources and improved management and control. The centralized SDN controller has the capability to efficiently control the flow of traffic by means of sending transmission commands to switches (via open flow) to command the appropriate operation of switches in terms of packet forwarding and processing. Finally, network administrators can have the ability to control the entire network using network applications through the technique of using a centralized controller as compared to managing and configuring each and every individual switch. In Figure. 3.3 there are 3 main layers that define the SDN architecture, namely, application layer, control plane and data forwarding plane. These layers are further described in details.

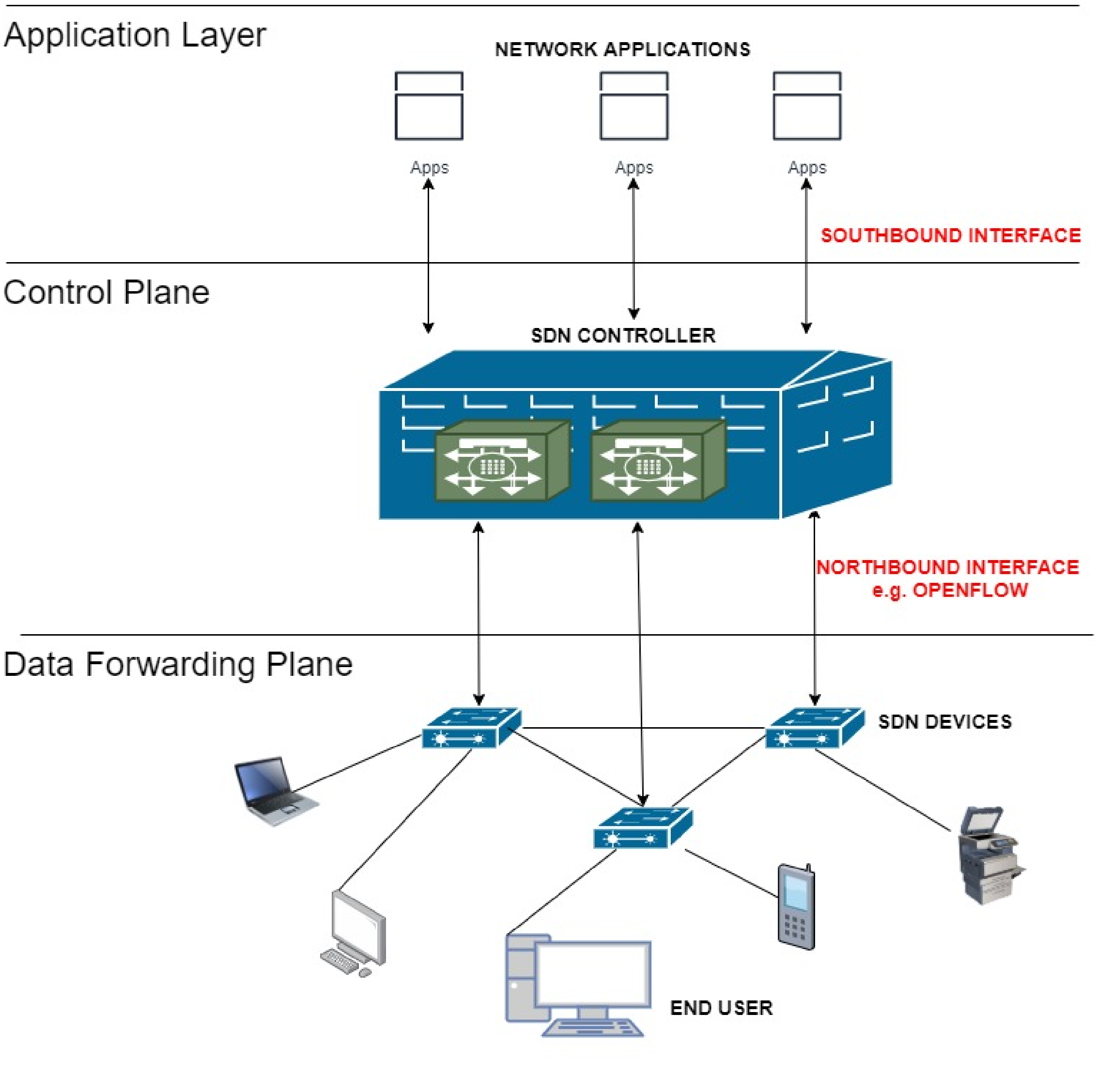


Figure 3. 3 Network architecture for Software Defined Networking (SDN)

A. Application Layer

It comprises of smart applications for end user interface which make use of the SDN-based communication and SDN-based network services such as quality of service, traffic engineering, access management and control, load balancing, and security as a service as well as the other virtualized network function services [29]. These end-user network applications are significant to configure and manage the SDN-based devices by communicating their requests via the upper interface (southbound) and lower interface (northbound).

B. Data Forwarding Plane

It comprises basically of SDN devices, it can be a physical device or a virtual device that is responsible for data forwarding and packet switching. Basically an SDN-based device is made up of a north bound interface which is responsible for transferring commands to and from the controller, a separation layer for device abstraction, and lastly a packet forwarding component which forwards packets of data based on the SDN controller command [30].

C. Control Plane

This plane comprises of a combination of connected controllers that communicate with upper applications and underlying devices through the southbound and northbound interfaces respectively. Basically the SDN controller controls all the devices through OpenFlow protocol and then the controllers communicates with other connected controllers on the side bound interface to achieve an uninterrupted global view of the entire network architecture. The SDN controller is obviously the intelligence of the entire network, therefore it is significant to go into further details on how this centralized controller operates.

3.2.1.1 The SDN Centralized Controller

An underlying SDN-based device observes the SDN centralized controller/manager as more of a controller software that institutes and configures the SDN-based device to a certain operational mode by programming to and from the memory of the device. Alternatively, the centralized controller can be viewed as intelligent software that controls and manages the underlying network, and efficiently utilize all the underlying network resources and infrastructure [31]. Figure 3.4 illustrates the main building blocks of a centralized SDN controller, including the principal operational units/modules, application modules, and interfaces. These building blocks are described as follows [32]:

A. The Controller Block

The controller unit is designed as the main block in which all devices such as switches are connected to it, and it provides the ability to send and receive requests from these switches.

B. Packet Processing Block

This block is designed to process packets based on their headers. This packet processing approach is achieved through network protocols in which each packet is modelled with different addresses corresponding to the source and intended destination and also the type of message transmitted.

C. Management of Device

This is the device manager unit which is responsible for managing network devices whereby each device is modelled with different network addresses such as switch address and switch port address in which it is connected to.

D. Topology Manager

This unit is responsible for identifying any alterations in the topology of the network. It also maintains the frequently updated network topology table and sends the new updates regarding the network topology in the application layer, so that the end user can observe and monitor the network topology at any time for any change.

E. Routing

The routing manager initiates an efficient route between the source and the destination.

F. Service Abstraction

The abstraction service layer is in control of the addition of new device protocols to the controller.

G. Interface Management

This is the unit responsible for the process of managing interfaces in the network which is generally given as web interfaces for entry to the controller. Additionally, these interfaces are faster techniques used to access specific functional operations which are provided by the controller.

H. Implementation of OpenFlow

Every controller must have an OpenFlow module that is responsible for maintaining OpenFlow protocols for messaging and routing.

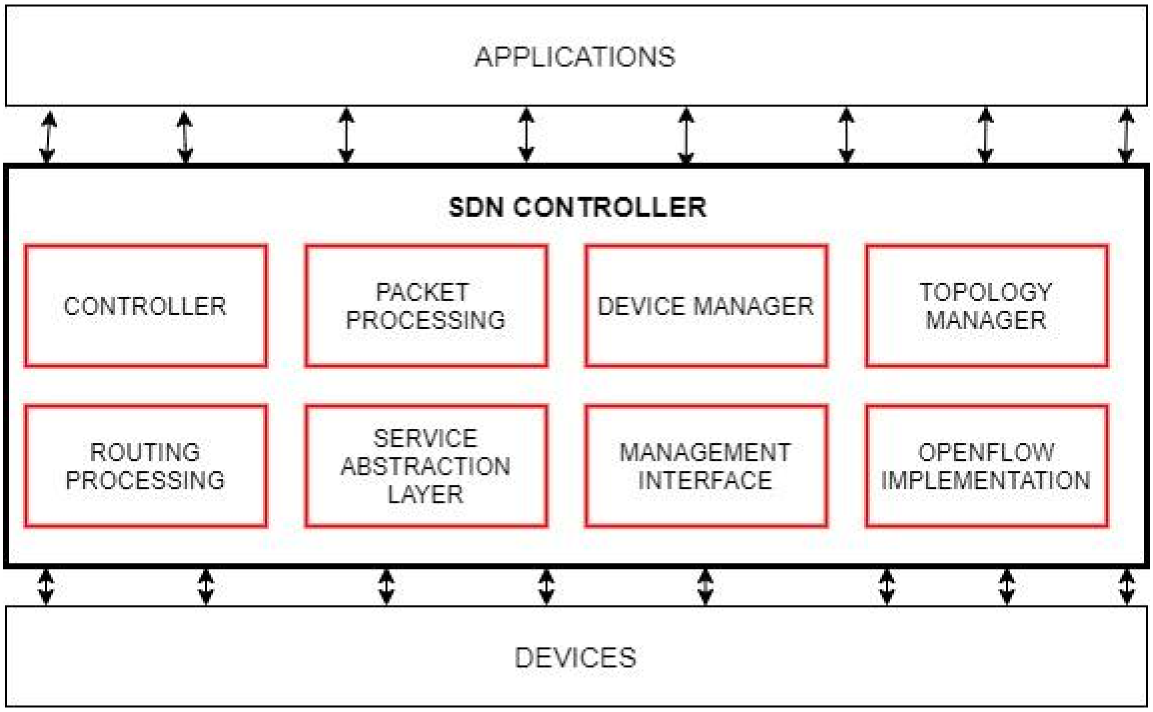


Figure 3. 4 SDN Controller core components [32]

There are still some issues with SDN which may need deeper research, such as the drawback of excessive requests to the SDN controller which in turn means extensive processing by the controller which may result in increased processing time hence increased latency as well as increased power consumption. With the controller being responsible for all the processing of information, then congestion problems might occur during busy hours with too many requests to the controller. This problem can be solved using load-balancing techniques efficiently or maybe having distributed controllers to share the load. One technique which can be adopted is the load-balancing route management framework. The load-balancing route management framework (LBRM) is responsible for continuously tracking the network traffic load and evaluate a load deviation framework to monitor if some of link routes are heavily loaded or if there is any possibility of congested routes. If so, then alternative paths are taken and the load is evenly distributed and shared to achieve load balancing, this is achieved through the use of adaptive route modification (ARM) mechanism [33]. Moreover, the use of techniques such as dynamic information polling (DIP) is also essential due to the fact that every network device (switches) can be inquired in-advance about their traffic load status in an attempt to reduce overhead messages to the controller. The features of LBRM such as ARM mechanism provide advantages of a stable network in which unnecessary frequent change in routes can be avoided, as well as assisting in switches to keep well updated flow tables to avoid buffer overflow [30], and lastly, the DIP mechanism is essential for adjusting the query period therefore resulting in reduced overhead to the controller while maintaining updated and precise information of switches.

3.2.2 Network Functions Virtualization (NFV)

The introduction of SDN in the backbone network introduces new possibilities such as virtualization and optimization. NFV does not necessarily require SDN to be implemented, however, SDN and NFV work side by side. SDN is more of an enabling technology for NFV, which allows for simple implementation of the network overlay model. Future mobile network designs are based on energy efficiency, scalability and versatility which could be achieved by SDN/NFV. In order to effectively achieve future network demands, there is a necessity to implement hardware network functions as virtualized network functions (software-based functions) using the Network Functions Virtualization paradigm. Network designers and vendors use NFV to implement network functions (which were previously hardware-based functions) to software-based functions which are referred to as virtual network functions (VNFs). These VNFs are deployed to cloud servers or data centers instead of specialized hardware [34]. NFV can play a major role in the backbone network and in the access network for current and future mobile networks. Firstly, we discuss the significance of NFV in the backbone core network to achieve energy efficiency. With NFV in place, EPC specialized functions such as the serving and packet data gateway and the mobility management entity can be virtualized and pooled centrally for easy management and control using SDN. On the access network, the baseband processing unit functions, including radio resource control, radio link control and medium access control functions can be virtualized as well. If a core backbone network architecture based on SDN and NFV was considered, then NFV takes advantage of the of the concept of virtualization by basically extracting functions away from dedicated hardware and deploying VNFs on commodity servers in the cloud infrastructure.

The current LTE-based core backbone network which is generally referred to as the Evolved Packet Core (EPC) as mentioned previously is illustrated in Figure 3.5, it comprises of various network functions which are implemented for dedicated operations that form part of the mobile core network. These functions are essential for any mobile core network because they play different role for example the MME and HSS are responsible for handling the control plane, while on the other hand the S-GW and P-GW are responsible for handling both the forwarding plane and the control plane. In the EPC architecture, the control plane functions, and the forwarding plane functions are deployed as physical hardware functions [34]. However, future networks such as 5G and beyond can leverage on SDN and NFV in the mobile core backbone network which will result in minimizing power consumption and reducing cost by extracting both the control plane functions (MME, HSS) and the data plane function (P-GW, S-GW) from their proprietary hardware to run on the cloud infrastructure. This approach is illustrated in Figure 3.6, where the architecture of the mobile core backbone network is now based on SDN and NFV. This architecture means the gateway's control plane and the data plane processing is now running on cloud data center as VNFs. With the centralization introduced by SDN and the virtualization introduced by NFV, all the EPC network functions are pooled as software running on general cloud servers. Thus, this results in using switches only for transport purposes, hence simple forwarding switches can be used in the core because there are no functions implemented on forwarding switches. Therefore, with all these mentioned advantages, operators can extremely benefit from this approach because of reducing CAPEX and OPEX in a sense that VNFs are used in this approach while also simple and cheap transport switches can also be used in the network without any consequence.

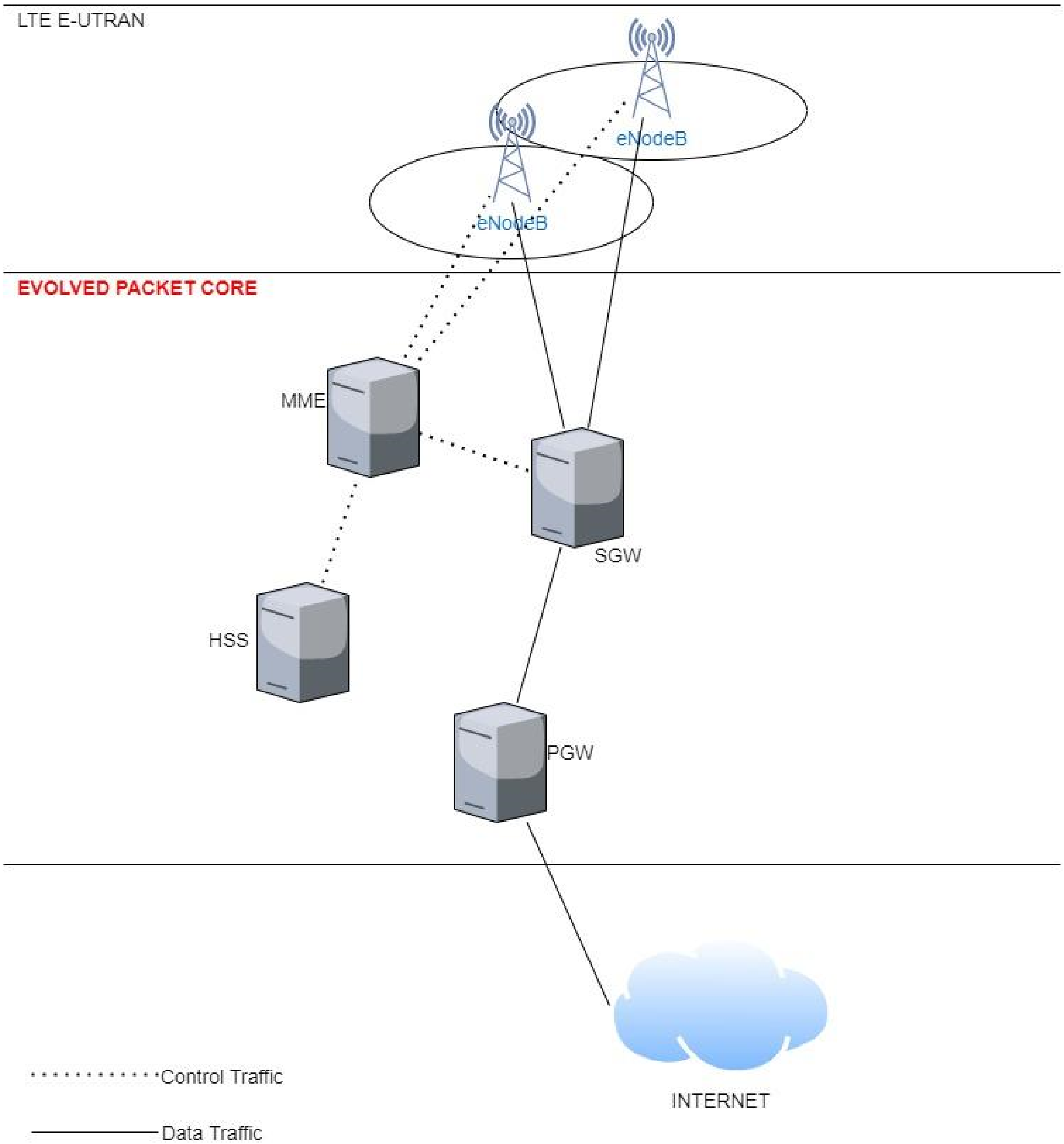


Figure 3. 5 LTE-based backbone core architecture (EPC)

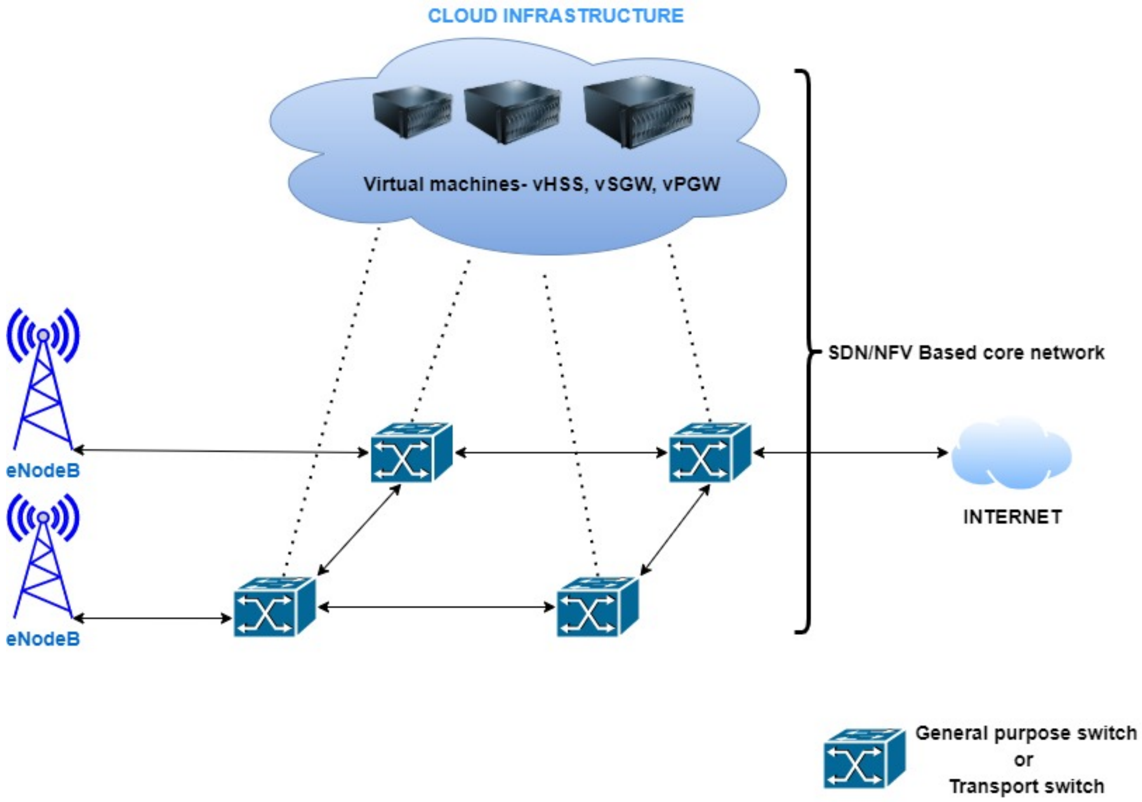


Figure 3. 6 SDN and NFV mobile core architecture.

The proposed core backbone architecture on Figure 3.6 introduces virtual machines i.e., virtual-based MME (V-MME), virtual-based HSS (V-HSS), virtual-based S-GW (VS-GW) and virtual-based P-GW (VP-GW) as compared to the LTE core network in Figure 3.4. The deployment of virtual machines can significantly reduce the energy consumed of the core network because of the exclusion of dedicated hardware, furthermore, virtual machines play a significant role in reducing OPEX and CAPEX hence virtual machines does not require any service operation but only software updates are required to adapt with the changing network demands. Finally, the proposed architecture leverages SDN capabilities to centralize the controller which results in using inexpensive switches (transport switches), which are only used for forwarding data with no processing or intelligence inside the switches.

3.2.3 Energy-Aware Routing Approaches for Efficient Core Networking

The aforementioned paradigms can be adopted in the backbone network but the effort can be futile if the overall operation of the network is not optimized correctly. Normally network components in the backbone core network are designed to support worse case traffic, however, this is not efficient if the traffic is at minimum levels because traffic levels vary from time to time, and therefore during low traffic periods huge amount of energy is used unwisely and wasted because some of the devices during that time will be idling without any functions being performed [36]. This issue has motivated for an efficient and intelligent way of networking and operation in the backbone network, and this will be achieved by intelligently switching off idling devices during low traffic periods, and this motivates for restructuring of routing schemes to be intelligently optimized to cater for dynamic network demands [37]. Therefore, routing schemes can be subdivided into three groups i.e., the energy aware routing, energy aware networking, and load adaptive operation.

3.2.3.1 Energy-Aware Routing

This technique can be viewed as intelligent approach of networking, whereby the nodes (switches) in the backbone network take an energy efficient route or in simple terms take the close route from source to destination. Energy aware routing creates new advantages whereby an energy aware platform is possible, this is because re-routing, traffic grooming and topology adaptation results in turning off devices that are not be used with the help of the SDN controller which can monitor all unused devices in the network. This is made possible by the implementation of SDN in the backbone core network which opens new possibilities of networking. This technique basically focuses on the reduction of the number of light paths.

3.2.3.2 Energy-Aware Networking

This is also an energy efficient approach of networking and can be similar to energy-aware routing but differs on a sense that it is based on previously frequent occurring traffic demands which are monitored and saved by the SDN controller. This technique of networking can be typically regarded as nodes going into sleep mode operation based on their daily traffic variation statistics. With SDN paradigm, the daily traffic stats can be collected and stored in the database, and then used as a reference to switch off all unused devices. This technique thrives to shut down low power devices as possible. Several approaches may be implemented to switch off the core network devices based on whether a node is completely redundant and also based on whether the traffic passing through a node falls below a specific threshold. Figure 3.7 illustrates the approach of energy aware networking and energy aware routing.

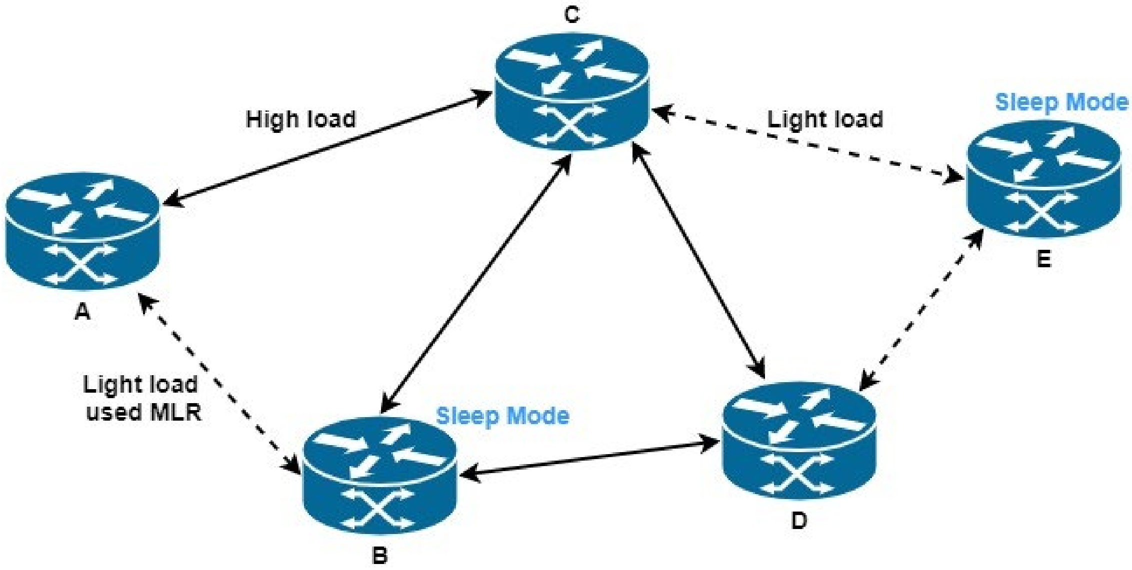


Figure 3. 7 Energy aware networking and routing

3.2.3.3 Load Adaptive Operation

In a network, high-capacity devices have proven to consume more power per bit of transmitted traffic, therefore, to achieve lower power consumption, it is a smart solution to deploy the least number of network devices with aggregated capacity that can accommodate the exact traffic demand. Adaptation to the heterogeneity of current networks as well as traffic demand has been achieved so far by the using mixed-line-rate (MLR) solutions [39] [40]. MLR can adjust the network bandwidth to that required by individual lines and thus lead to appreciable energy savings as opposed to networks that only support single line rate (SLR) [41]. A network that supports MLR is generally called an elastic network because of the improved elasticity of the network as compared to SLR. For example, if node A and B in Figure 3.7 are lightly loaded, then their capacity is adjusted such that they continue to exist and operate in the network but at a much lower speed or capacity hence lower energy consumption. Traditionally, most optical core backbone networks operate on single line rate such as 100 Gbps despite different capacity demands that actually exist between nodes [42], therefore it is clear that on low traffic loads using a SLR of 100 Gbps is extremely inefficient and shows poor utilization of resources which then results in unnecessary usage of energy. The approach of using MLR appears to be a significant step in the right direction as far as power consumption is concerned in current and future networks. Despite directly lowering energy consumption due to their adaptability to variable light path patterns, MLR also improves the overall spectral efficiency of links and thus impact positively on the general energy efficiency of the entire core backbone network

3.3 Energy Efficient design approaches for RAN.

The next step herein would be to achieve energy efficiency in the design and operation of a RAN architecture. Demands are high for current and future networks such as 5G networks to accommodate massive amount of data, continuous connectivity i.e., anytime anywhere connectivity, and this is due to a massive increase to the number of devices coming online which includes Internet of Things (IoT). With more bandwidth hungry devices coming online, this further implies a drastic increase in energy consumption. . Noting that the RAN i.e., the base stations, consumes approximately 60% to 70% of the energy within the network, while also contributing a large percentage to the carbon emission as well [45]. This has motivated the introduction of green routing dynamisms directed at reducing energy consumption in RAN components and curbing carbon emissions. Therefore, the RAN is the main focal point towards achieving energy efficiency in mobile networks. According to [46], base station power consumption at peak loads can be represented in Figure 3.8, where the building blocks of a base station are represented and their power consumption.

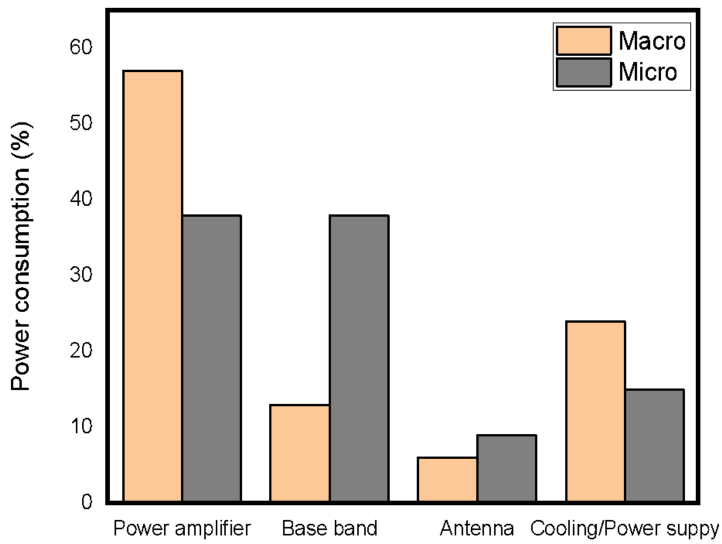


Figure 3. 8 Base Station power consumption.

In Figure 3.8 above it can be observed that the power amplifier section of the base station is consuming more power whether it's a macro or micro base station as compared to other sections i.e. base band, antenna and power supply sections. The overall consumption of power from all the sections of the base station proves that this issue of energy consumption cannot be neglected any further. Despite the positive environmental effect of a reduced power consumption in telecommunication networks, it is also predicted that network operators can mutually benefit from reduced power consumption in a sense that there is a possibility of reducing OPEX.

3.3.1 Power Consumption of a Base Station.

Energy efficiency/power consumption of a base station is a significant parameter to observe for any given wireless access network. Thus, for any given base station, the energy efficiency can be represented as a power consumption required to cover a specific area or cell coverage, which can be measured in W/m^2. Therefore, Energy Efficiency (EE) of a base station can be mathematically defined as the power consumption (PC) per covered area.

$$\mathrm{EE} = PC_{area} \tag{3.1}$$

$$PC_{area} = \frac{P_{el}}{\pi R^2} \tag{3.2}$$

Where P_{el} is the electrical energy consumption of the base station during power up, and R is the coverage range of a given base station. From equation 3.2, Assuming that the PC_{area} is low, then the base station is termed energy efficient. It can be observed that a macro-cell base station and a small-cell base station will consume different power because of the difference in coverage range. For example, assuming a small-cell coverage range is half that of a macro-cell coverage range then using equation 3.2, it is observed that 4 times the power consumption of a small-cell base station is required to cover a single macro-cell coverage, which denotes that in this case the small-cell base station is energy inefficient as shown below.

$$PC_{area} = \frac{P_{el}}{\pi R^2}$$

$$PC_{area} = \frac{P_{el}}{\pi (\frac{R}{2})^2}$$

$$PC_{area} = \frac{P_{el}}{\frac{\pi R^2}{4}}$$

$$PC_{area} = 4 * (\frac{P_{el}}{\pi R^2})$$

However, this does not mean small-cell base stations are always energy inefficient but rather it depends on the coverage range. For example, if a small area or short range needs to be covered, then deploying a macro-cell base station would mean a lot of power is wasted unnecessary because the coverage range is short thus a small-cell base station would be necessary to do the job at a much-reduced power.

Power efficiency can also be represented by the number of transmitted bits per joule of energy consumed [49].

$$Energy\ Efficiency\ (EE) = \frac{Transmitted\ bit}{Energy\ consumed} \quad (3.3)$$

$$EE = \frac{(U_{BS}/100) * Bit_S * RB_S * RB_n}{P_{avg}} \quad (3.4)$$

Where:

- $U_{BS}/100$ represents the average utilization of a base station in percentage form.
- Bit_s represents bits per symbol.
- RB_s represents the total available resource blocks.
- RB_n represents the total symbols in a block.
- P_{avg} represents base station power consumption average.

Towards achieving an energy efficient RAN, energy efficient techniques are broken down into smaller segments at each node to achieve the overall RAN efficiency. These techniques are divided into different segments i.e., efficient wireless node re-design (from component level to module level), efficient wireless node operation, efficient transport (node-to-node interaction).

3.3.2 Efficient Wireless Node Re-Design

RAN Nodes or wireless nodes are typically base stations which primarily comprise of power supply modules, transceiver modules, and processing modules [50]. To achieve energy efficiency at the overall network, efficiency should be achieved at each node, furthermore, efficiency should be achieved at each component that build the node. Hence each module or component is vital to its own energy efficiency since component inefficiencies add up to the overall node inefficiency, and consequently adding up to the overall network inefficiency. Thus the first step in re-designing an efficient node is to use efficient components. This means for processing, only logic devices should be used, mainly because logic devices are usually CMOS-based devices which have lower static power consumption [51]. The issue with logic devices is the heating up during switching operations, which in turns requires some cooling operations to overcome short term inefficiencies, but these cooling operations further increase operational power consumption as shown in Figure 3.8. Low-rate processors can reduce the power since less energy is used up for computation, meaning less energy used for cooling. This is only possible, if high-rate processing is centralized and migrated to the cloud infrastructure where virtualization is possible. This means functions with major processing such as baseband processing can be pooled and centralized in the cloud, this centralization approach of baseband units (BBU) is of vital importance in reducing energy of the overall network, rather than having baseband units on each node, this approach is also known as Centralized RAN (CRAN).

I. Cloud/Centralized RAN (CRAN)

The RAN of the previous generations such as 2G was designed in such a way that all the frequency-based functions (i.e., conversions and amplification) and baseband processing was performed at the wireless node site as mentioned in chapter 2. As the generation of mobile networks continue to improve, that meant more demands of data which then made it complex to perform all these functions in one location, hence the introduction of Distributed RAN in 3G and 4G. 5G network on the other hand introduces a new RAN concept which has been designed to accommodate massive data transmission required for 5G while minimizing power and cost at the same time, this concept is called the Cloud-RAN (CRAN), or Centralized RAN as other authors refer to it. CRAN refers to the concept of merging all data to a centralized location; basically C-RAN in wireless networks can be viewed as a network architecture where baseband infrastructure/resources are consolidated to a central location, so that resources are shared amongst base stations. In this section herein CAPEX can be viewed as expenditure related to network construction which includes network planning, RF hardware design, baseband hardware design, installation and software licensing, while OPEX on the other hand relates to cost that are required for network operation such as site rentals, electricity, operation and maintenance such as upgrades. Generally, a CRAN architecture comprise of three main components i.e., the nodes (Remote Radio Head), centralized baseband unit (BBU) pool, and an optical network which provide connectivity between the RRH and the BBU pool which is generally referred to as a front-haul network. The main advantage of this concept is that the BBU pool has a number of software defined BBUs with centralized processors to optimize radio resource, which is why this technique is considered energy efficient and also cost efficient. Since 5G is considered a heterogeneous network with various deployment of small cell technology together with macro cells, then heterogeneous CRAN (H-CRAN) is applicable for 5G networks. H-CRAN refers to the merging of CRAN together with the heterogeneity of networks such as 5G to achieve major benefits for resource allocation. This is achieved because the heterogeneity of 5G is created to provide high data rates capacity and improved QoS, while the introduction of CRAN will leverage benefits of energy efficiency as it resides all processing of digital data in the BBU pool, which basically mean H-CRAN is more power efficient while at the same time providing increased throughput of data rates. It should be noted that with H-CRAN there could be significant reduction in energy consumption because the BBUs are placed in data centers, which means the wireless nodes are no longer responsible for processing and therefore consume minimum power as compared to previous generations. Figure 3.9 represents an architecture for H-CRAN.

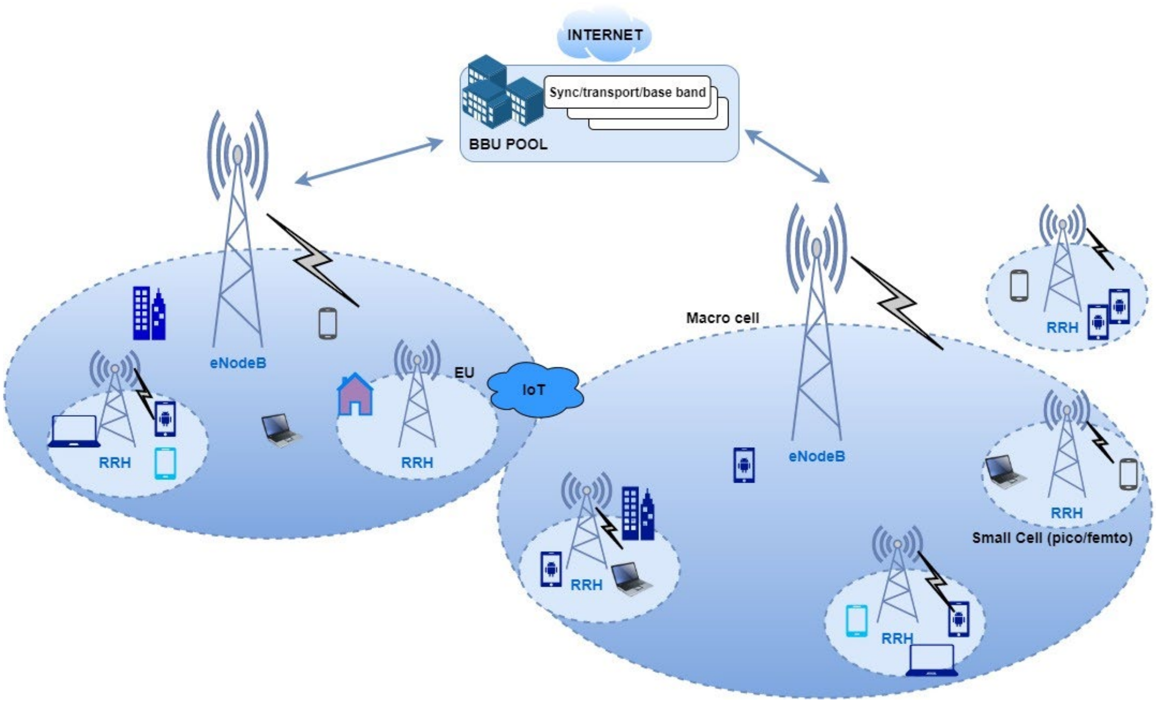


Figure 3. 9 Typical H-CRAN architecture.

There are many advantages of H-CRAN, both in small cells and macro cells, this is due to the efficient optimization of BBUs and reduction of cost to base station deployment. A few advantages of H-CRAN are mentioned below.

- **Energy and cost savings** – The deployment of H-CRAN architecture can significantly reduce power consumption, consequently reducing cost. Most of the energy in a cellular network is drawn by power amplifiers, wireless nodes (RRH), and cooling systems of nodes and baseband processing, thus with H-CRAN, baseband processing of several wireless nodes is centralized and co-located in one area, hence this technique can decrease power consumption and cost since electricity is used only at one point rather than at each and every wireless node. Furthermore, BBU pooling allows for switching off on some BBUs during low traffic periods since every BBU is co-located in one area. Finally, maintenance work performed on distributed site locations can be narrowed down to a central location where equipment are co-located in one site, which will further decrease OPEX.

- **Decreased delays and improved throughput** – H-CRAN architecture can also decrease delays because of pooling BBUs in a more convenient location, and processing of signals from many base stations can be done on the BBU pool resulting in reduced processing and transmission delays, as a consequence throughput will be improved.
- **Adaptability and scalability to various traffic** – Traffic varies during the day and night, which means base station utilization varies, thus an adaptable network architecture is essential for achieving efficient utilization of resources. CRAN baseband processing of various cells is achieved and controlled in a centralized BBU pool hence the overall utilization rate of the network can be improved. Moreover, It is expected that the overall baseband capacity of a BBU pool will be smaller than the total of individual baseband capacities of single base stations [53].

II. Mobile Edge Computing (MEC)

There are other techniques which can greatly improve the power efficiency of current and future wireless networks, such as Mobile Edge Computing (MEC), this is another technique which can be adopted for energy efficiency in the current and future generation of cellular networks. MEC basically provides cloud or fog computing capabilities close to the users at network edge, within the RAN. This means mobile devices have the ability to offload their tasks on MEC servers which are located closer rather than using servers on the backbone core network which are generally located far [54]. This approach not only reduces latency but also offers computational offloading. Computational offloading provides an edge of power consumption, transmission time, and overall performance [55]. Figure 3.10 illustrates how the base stations can include MEC servers on their architecture.

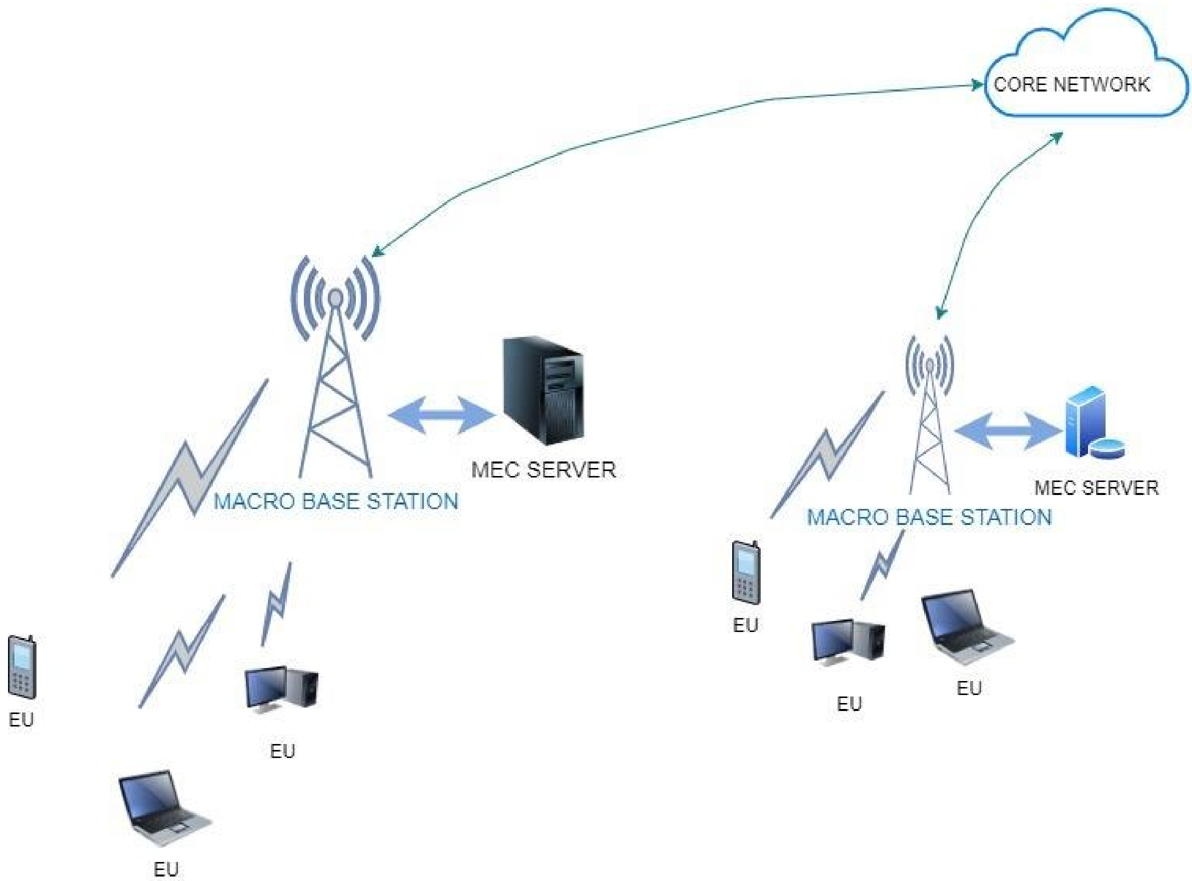


Figure 3. 10 Mobile Edge Computing offloading architecture [56].

III. Beamforming with Massive Multiple Input, Multiple Output (MIMO) Antenna.

Beamforming is another efficient wireless technique that focuses a radio signal directly towards the intended device, rather than having the signal spreading in all directions from a broadcast antenna. This is a technique used to create pattern of an array antenna by combining constructively the weights of the signals towards the intended signal and nulling the pattern towards the direction of unwanted signals. This method of focusing a signal directly to a specific direction allows for higher signal quality delivered to the receiver without needing to boost broadcast power, which is a huge advantage since boosting power requires a power amplifier and as observed in Figure 3.8 power amplifiers are huge consumers of energy. Beamforming is a key tool to exploit the potential of massive MIMO systems. Massive MIMO wireless communication refers to the mobile base stations which has a massive number of antennas. This approach has proven to improve the overall spectral and energy efficiency [57]. Massive MIMO replaces conventional array antennas which use expensive hardware but consist of few antennas with hundreds of small antennas that use inexpensive amplifier circuitry. These MIMO antennas are considered energy efficient because they have been proven to reduce the radiated power, while keeping the information rate unaltered [58]. The advantages of massive MIMO antennas can be leveraged using beamforming. This can be regarded as a technique for simultaneously transmitting and receiving multiple signals of data on a common RF channel through multi-path propagation. Massive MIMO is a significant technology in 5G networks and future networks because it can accommodate a very huge number of antennas typically over 100 antennas. Since 5G network is expected to operate using millimetre wave (mmWave) frequency in order to achieve higher data rates, but mmWave suffers from propagation loss, however, with the ability of massive MIMO to provide higher beamforming gain which can then compensate for larger propagation loss in higher frequencies. Additionally, massive MIMO also offers re-use of the same frequency band for many users [59] [60]. Beamforming together with massive MIMO can be employed at both the transmitter and receiver ends in order to achieve appropriate feeding for the antennas to steer the beams towards the desired directions. With this technique on current and future networks, energy can be highly conserved and used efficiently. Massive MIMO with beamforming is illustrated in Figure 3.11.

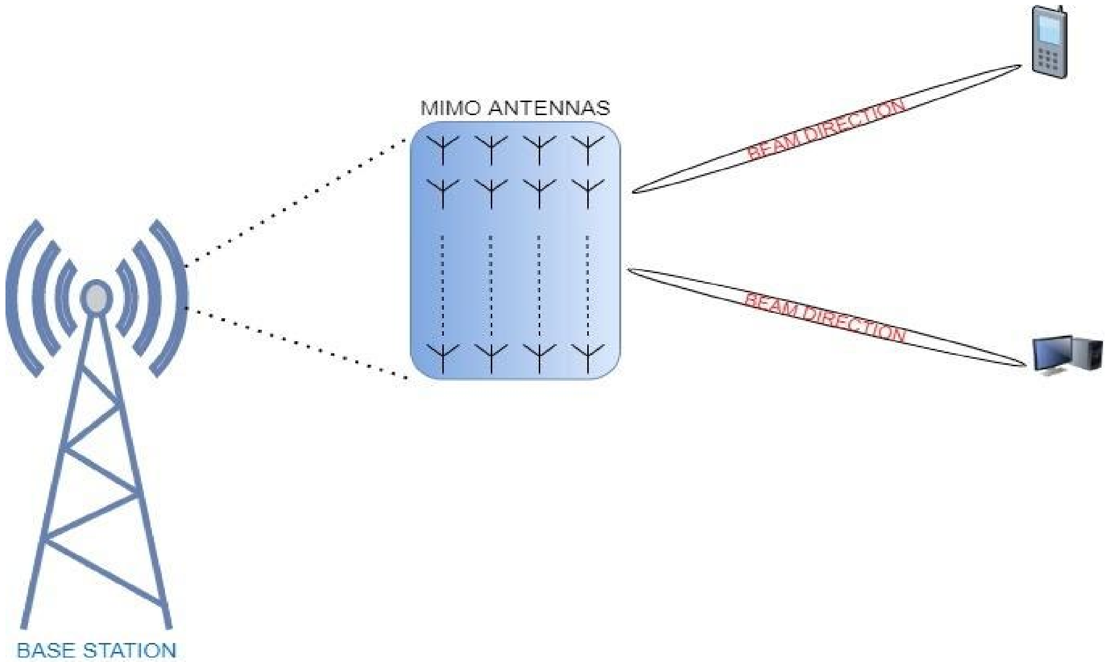


Figure 3. 11 Massive MIMO with beamforming

IV. Device-to-Device Communication (D2D)

This is a new paradigm to be employed in the next generation of mobile network, while in conventional networks, mobile users are not permitted to directly communicate with each other, but with the introduction of device-to-device communication this can now be allowed whereby several co-located devices or devices in close proximity have the ability to directly communicate using a mobile frequency rather than the signal going through the wireless node. This technique promises to be highly energy efficient since direct communication between nearby devices may occur at a much decreased transmit power than that required for transmitting through a base station that can be located far away. This approach creates benefits such as low latency because the signal goes directly to the user equipment as compared to going via the base station, and additionally this technique helps on offloading base stations. A device with an acceptable internet connectivity can act as a relay to which data is offloaded from the base station and which other devices can download data using D2D links, furthermore, devices that lack processing power may also offload computational-heavy activities to closer and more capable devices using D2D links [61]. Finally, this technique can also play significant role in coverage extension, for example when a mobile user is at the edge of cell coverage then the user may receive poor signal quality while connecting to the base station and if there is another mobile user close to the base station and having a fairly good link to the base station may act as a relay for out of reach devices. Figure 3.12 represents a cellular communication and D2D communication including D2D relay.

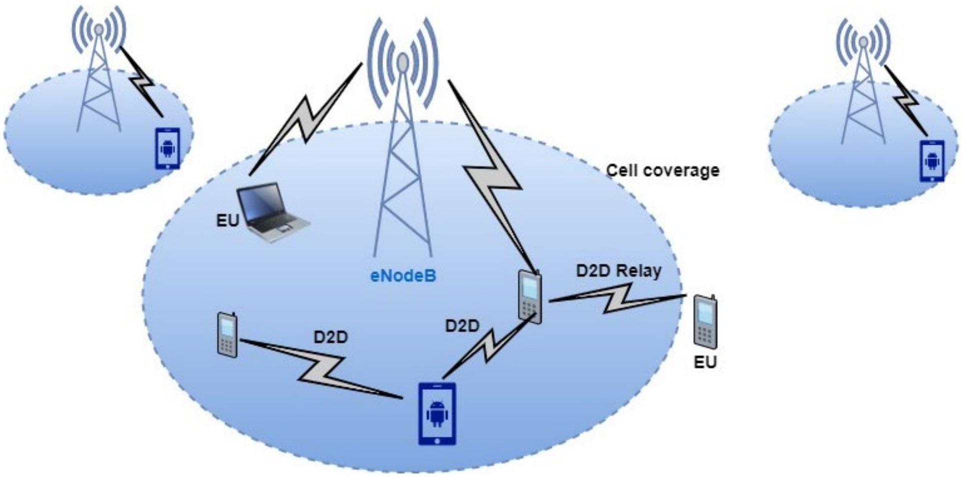


Figure 3. 12 D2D communication

3.3.3 Efficient Wireless Node Operation

As mentioned before that previous generations were only focused on improved quality of service (QoS) and coverage, not so much on energy efficient operation of the network which in turn resulted in an "always on" operation of the network. To achieve an efficient node operation, network devices should not always be on because practically they are not always needed to be in use all at once, except in those special operations where services and applications don't need any disturbance or interruptions. In this section, the focus is based on the operation of the wireless node and considers how to manage its power during operation based on the demands of traffic at any given time.

I. Node Sleep

This approach of having nodes into sleep mode when they are inactive has proven to reduce inefficiencies, and which has led to this technique being adopted in the operation of current 5G network and beyond 5G. This technique may suffer negative impact such as delays which can degrade the QoS, but in terms of power savings, this approach promises to save a lot of power. The activation of sleep mode operation is based on load demands or time allocation during off-peak times. The node can either be triggered using a passive start up receiver [62] [63], or arranged to wake up at a specific time, which means the wireless nodes are scheduled for a certain time allocation for communication and afterwards they are permitted to sleep to further become triggered at their allocated time [64]. This technique gives advantages because the transceiver only wakes up when communication is necessary, while on previous generations the node is always on even if traffic demands are low which then consumes unnecessary power. This is important because if the wireless node is on sleep mode then the power amplifier shuts down and hence the cooling power of the power amplifier shuts down as well, noting that the power amplifier is the highest power consumer. Since 5G network is a heterogeneous network with a multi-tier coverage, that simply means a denser deployment of pico, femto and micro cells on the network, therefore, it is expected that different mobile cell sizes will occupy different load demands at each time, so in order to manage power in all these diverse cells, it is thus imperative to only activate a base stations if it is necessary otherwise it should be deactivated (sleep mode). Sleep modes can be divided into two categories namely deep sleep operation mode and micro sleep operation mode. Micro sleep operational mode of a base station can suspend its transmission for order of milliseconds hence can be prompted to become alive almost immediately, while in deep sleep operational mode the base station can turn off for longer time periods [65].

To maintain an acceptable QoS, then node sleep mode will depend on the cell type i.e., micro, macro, pico or femto, and also depend on the amount of traffic at a given node or base station. To elaborate on this further, if a macro cell base station is on sleep mode, then the capacity of that cell will be reduced. However, for a small cell i.e., femto cell and pico cell, when the sleep mode is activated then the overall cell traffic is transferred to macro cell base station, which means base stations for small cells play a major role in offloading the macro base station. Algorithm 1 proposes how the node sleep operation can be achieved, whereby the sleep mode operation is activated when the base station traffic is less than its capacity, and when the base station traffic is greater than its capacity then power-up is activated. The logic behind this operation is based on keeping the traffic load at each base station within its capacity.

Algorithm 1 Sleep mode operation for a base station

```
for all base station do
if base station type = Small cell then
{
if base station traffic < base station capacitythen
Sleep mode operation for small cell activated
}
else if base station traffic > base station capacitythen
Normal operation for small cell activated
else if base station type = Macro cell then
{
if base station traffic < base station capacitythen
Sleep mode operation for macro cell activated
else if base station traffic > base station capacitythen
Normal operation for macro cell activated
}
End
```

Nodes can go into sleep mode when traffic is low, but for those nodes which are awake it is important to efficiently conserve power while managing their activities using techniques based on adaptive node and component operation. This technique offers intelligent operation of nodes and components based on the loading effect. Variations in dynamic traffic is continuously tracked to observe when to shut a component down to manage energy efficiently, for example, in high-power sensor applications, adaptive sensing techniques can be used to monitor and trigger applications only when it is necessary [66] [67]. Finally, processors can be used efficiently using Dynamic Power Management (DPM) which is an approach that uses logic device optimization [68], whereby the DPM shuts the processor down to save power when computation is not required.

II. Cognitive Radio Networks

This is a promising technology which has the potential to satisfy the strict spectrum requirements for a network such as 5G. As the name implies, cognitive radio has cognitive capabilities and reconfiguration abilities during operation. Cognitive radio is basically viewed as a smart radio that can change and adapt its transmission parameters (power, modulation, frequency spectrum etc.) with the ability of re-arranging hardware as well as software according to the dynamic environmental radio conditions in which it operates. The basic functions of a cognitive radio may include sensing, analyzing and managing the spectrum, mobility management, and lastly spectrum allocation and sharing. This is illustrated in Figure 3.13 as a cognitive cycle. The way cognitive radios operate is based on cognitive operation, through which before a cognitive radio can go to its transmission operation mode, firstly cognitive radio devices must be aware of the transmitted waveform, radio frequency spectrum, communication protocol, geographical information, user demands, available local resources and security policy [69]. With all this gathered information by cognitive radio devices, then it is now more capable to dynamically change (reconfigure) the transmission parameters accordingly to satisfy the demands and achieve optimal performance [69]. Finally, cognitive radio has the ability to sense and based on the sensed information, it can neglect other occupied user frequency band hence avoiding unnecessary interference between users.

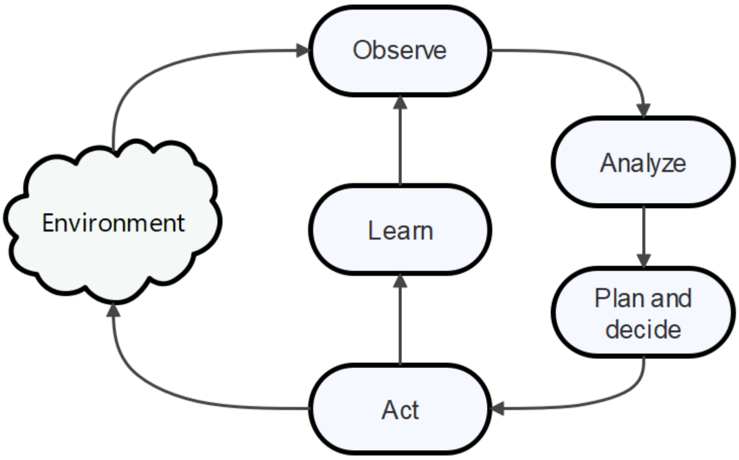


Figure 3. 13 Cognitive cycle

3.4 Summary Chapter Conclusions

In this chapter we discussed energy efficiency approaches for WTNs, which includes the RAN and the backbone core network. On the backbone core network, we proposed a flexible/elastic architecture which is based on SDN and NFV paradigms, which basically migrates core network functional entities to cloud infrastructure as compared to previous core networks (EPC) whereby network core functional entities have dedicated hardware. Furthermore, we discussed strategies of energy aware operation in the backbone core. Finally, we discussed energy efficiency architectural design approaches for RAN, which includes strategies such as efficient wireless node re-design, efficient wireless node operation, and efficient wireless node to node interaction. Efficient wireless node re-design was based on H-CRAN architecture, MEC architecture, MIMO antenna, and D2D communication. Efficient wireless node operation was based on sleep mode operation of base stations, in which an algorithm was proposed for sleep mode operation of a base station depending on the amount of traffic.

4. Resource Allocation and Energy Efficiency in Elastic OTNs.

4.1 Introduction

The concept of network data traffic is significant in network redesign, because traffic is generated when users share and transfer information, which can peak at certain times or in contrast be light at certain times. Traffic growth increases exponentially year after year, thus it is imperative to restructure and redesign fixed access transport network to adapt with the uncertainty of traffic evolution. This day-by-day traffic increase has resulted in an urge to new improvements in the field of networking. Hence this frequent increase of data traffic has motivated for a more flexible, scalable, adaptable and resilience optical transport network based on orthogonal frequency division multiplexing (OFDM) which is significant for mobile network back and front-haul. The term “flexible” in optical networks basically means the network can adapt and adjust its resources in a favorable manner to support the variations in traffic demands. Additionally, the increase of power consumption in the ICT is a significant problem for operators, thus the approach of flexible OFDM based optical transport network also aims to reduce power consumption and efficiently use the available spectrum.

4.2 Analytical Power Consumption in OTNs

According to [20], the total energy consumption in an optical multilayer backbone network can be divided into four constituting sections, namely, Ethernet power, IP power, WDM power and OTN power. The total energy consumption is thus expressed as the sum of the constituting layers.

$$P_{total} = P_{IP} + P_{Ethernet} + P_{OTN} + P_{WDM} \tag{4.1}$$

Where,

$$P_{WDM} = P_{OXC} + P_{Amplifier} + P_{transponder} + P_{regeneration} \tag{4.2}$$

From the expression in equation 4.2, the energy consumption of a WDM optical network (P_{WDM}) is further expressed as the total power consumption of the optical components such as the OXC, EDFA amplifier, transponder and regeneration. The concept of designing an energy efficient optical transport network is achieved through the deployment of energy efficient optical components known as bandwidth variable OXC, and the bandwidth variable transponder.

4.3 WDM Single Line Rate

Conventional core optical backbone networks used to support and operate at a single line rate (SLR) as the name implies, this line rate was either 10/40/100 Gbps. As mentioned in Chapter 2, single line rate networks are extremely inefficient due to their poor utilization of resources and poor utilization of energy. This poor utilization of resources further results in an increased OPEX and CAPEX which is a huge drawback for network operators. For example, if a 90 G traffic exists in a network between two nodes as shown in Figure 4.1, an efficient and effective way to deal with this traffic demand would be to deploy a single 100 Gbps transponder rather than deploying five 20 Gbps transponders to achieve the exact same task. This would mean buying 5 transponders instead of only one and further consuming five times the power.

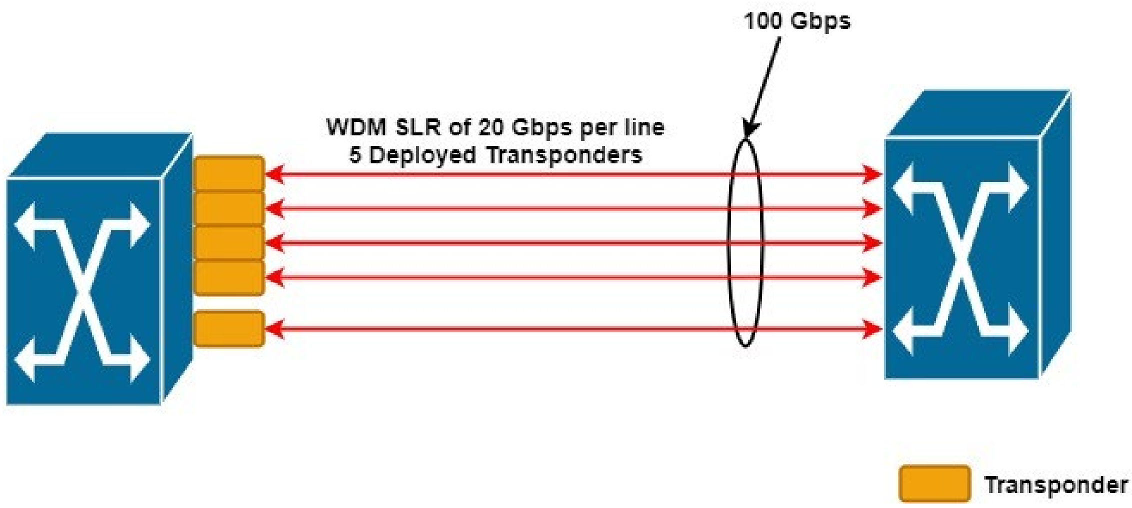


Figure 4. 1 Single line rate

4.4 WDM Mixed Line Rate

Mixed line rate (MLR) can be viewed as a milestone in transport networks as compared to SLR. Mixed line rate or multi-line rate as referred to by other authors, is basically a network employing various line rates which is clearly much more efficient than conventional SLR networks. In today's network, traffic is more heterogeneous, so the ability of MLR to incorporate lines with different data rates like 10/40/100 Gbps is clearly a huge improvement from SLR. Furthermore, different line rates use different modulation formats which are multiplexed into a single mode fiber as shown in Figure 4.2. To elaborate further, 10 Gbps data rates line usually uses On-Off Keying (OOK) modulation, 40 Gbps line rates use Quadrature Phase Shift Keying (QPSK) modulation and finally 100 Gbps line use Differential Quadrature Phase Shift Keying (DQPSK) modulation as shown in Table 4.1. The reason for opting to use PSK modulation for line rates above 40 Gbps rather than still using OOK is to avoid interference related issues, since OOK is prone to various impairments such as dispersion. The distance travelled by an optical signal before its BER and quality degrades to an unrecoverable level is known as a transmission reach, and the transmission reach of each modulation format is also shown in Table 4.1.

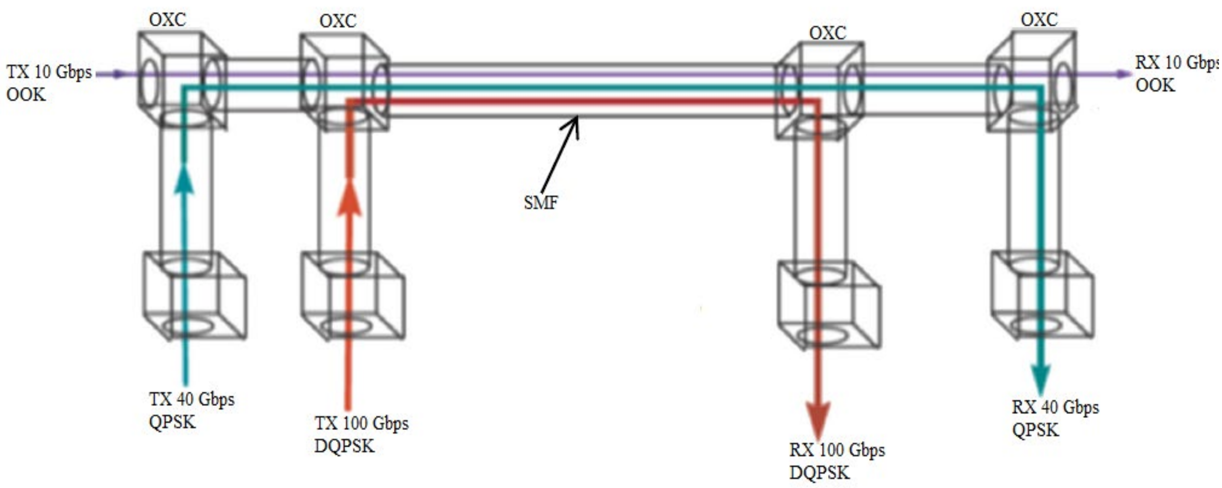


Figure 4. 2 Mixed line rate network [70]

Table 4. 1 WDM modulation formats and their respective transmission reach

Data rates	Modulation formats	Transmission reach
10 Gbps	OOK	2500 km
40 Gbps	QPSK	1500 km
100 Gbps	DQPSK	800 km

4.5 Elastic/Flexible-grid OFDM-based Optical Network

The is no signal conversion in transparent elastic optical networks because the is no re-generation but optical reach is limited even though the signal remains in the optical domain along the light path, while translucent elastic optical networks can increase the optical reach by including some regenerators where necessary. Therefore, despite the advancement made by MLR networks, there is still a necessity to have a flexible network to accommodate data rates of 100 Gbps and beyond. Flexible/elastic grid OFDM-based optical network is a new technology which promises an efficient and optimal operation of the network, it is elastic in a sense that it allows for a reconfigurable bandwidth transmission by allowing an adjustable number of subcarriers inhabiting a bandwidth equivalent to the appropriate subcarrier frequency slots, which is achieved by orthogonality between subcarriers. The principle used to maintain orthogonality among subcarriers is based on keeping subcarrier spacing equivalent to the baud rate, thus the subcarrier spacing maintains a fixed frequency and further instigating a minimum spectrum unit [72]. This type of network give rise to new possibilities of subcarriers overlapping in the spectrum domain (orthogonality), hence allowing for creation of super channels, which could not be achieved on conventional WDM networks due to the fixed-grid constraint. To understand the distinct advantage of flexible OFDM-based optical transport networks over WDM networks, firstly a brief overview on conventional WDM network must be described. WDM networks use a fixed grid (ITU grid) of 50 GHz with a maximum of 80 wavelengths in the C-band that support three different transmission rates of 10 Gbps, 40 Gbps and 100 Gbps using different modulation techniques namely, NRZ on-off keying, DQPSK and PDM-QPSK respectively. However, this technique is not flexible enough to accommodate network demands with different line rates, hence to improve its flexibility a network operation such as mixed line rates (MLR) was introduced on WDM networks. Additionally, to reduce the effect of cross-talk between adjacent channels in WDM, a guard band of 200 GHz was allocated [73]. The main issue with this approach was the allocation of the fixed ITU-grid with pre-defined spectral positions and bandwidth resulting in a rigid resource allocation hence that resulted in spectral resources being utilized inefficiently, as a consequence power can also be wasted because exact traffic demand may be much smaller than the allocated wavelength capacity. However, as mentioned above, an OFDM-based optical network is much more efficient in utilizing spectral resources because there is no ITU grid alignment instead subcarriers can overlap and allowing more compression on a bandwidth channel which results in higher data rates as compared to conventional WDM networks. Finally, OFDM introduces the advantage of using a DSP (Digital Signal Processor) in both the source end and destination end to achieve coherent detection, which results in the ability of using different subcarrier modulation techniques. An appropriate modulation format can be selected based on two conditions, i.e. the efficient spectral condition and the efficient energy condition. Figure 4.3 shows the flexibility and resource allocation of elastic bandwidth OFDM-based optical network as compared to conventional fixed grid WDM networks.

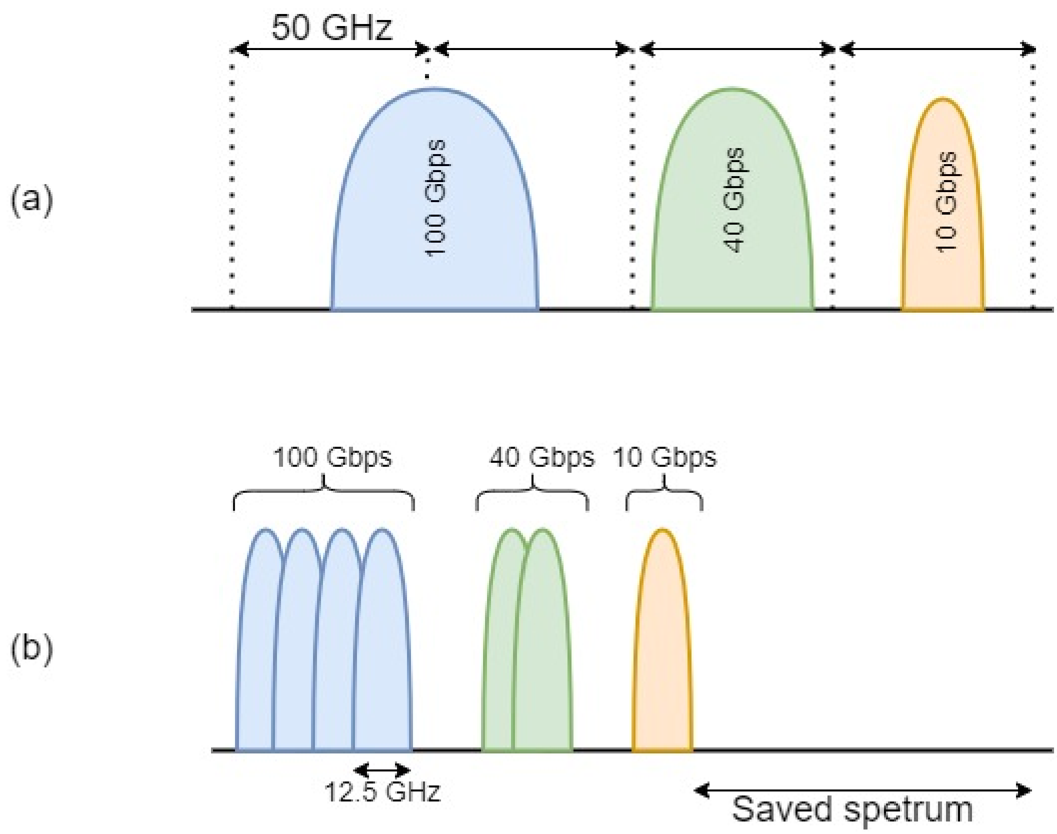


Figure 4. 3 (a) Conventional rigid WDM network (b) OFDM-based elastic network [74].

The significant distinct difference in the manner of deploying elastic bandwidth OFDM-based network as compared to conventional WDM network is mainly on deploying transponders which are capable dynamically changing bandwidth accordingly known as bandwidth variable transponders (BVT) and deploying optical cross connects which are also capable of changing bandwidth accordingly known as bandwidth variable optical cross connects (BV-OXC) as depicted in Figure 4.4.

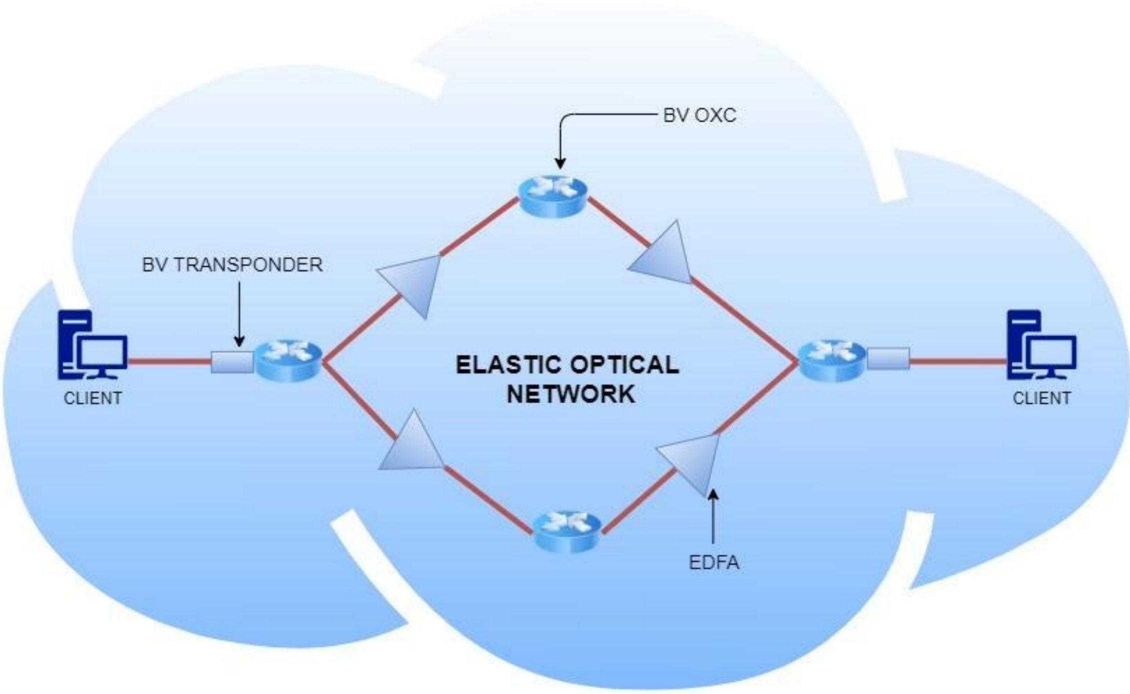


Figure 4. 4 General elastic optical path network with BV-OXC and BVT [75].

The network operates as follows, the BV transponder reconfigures a number of subcarriers (spectrum) and their modulation techniques according to the requested demand, which then means every BV-OXC along the path from source to destination sets the appropriately sized spectrum cross-connection which permits a suitable end-to-end optical path with bandwidth capacity which actually accommodates the spectrum of the path. Moreover, the BV-OXCs are distributing wavelength signals through switches at the output ports as shown in Figure 4.6. Likewise, bandwidth variable (BV) transponders are responsible for providing variable spectral domain and further enabling dynamic adjustment of capacity to cater for variable service demand. Furthermore, BV transponders can be adopted despite the use of different modulation techniques which is a huge upgrade from conventional WDM network, in which using different modulations formats can result in using different transponders or an entire upgrade in hardware [76]. As mentioned, this technique allows for elastic bandwidth transmission according to the requested capacity, so BV transponders can arrange the number of contiguous subcarriers and their modulation formats. A general block diagram representing a BV transponder is shown in Figure 4.5.

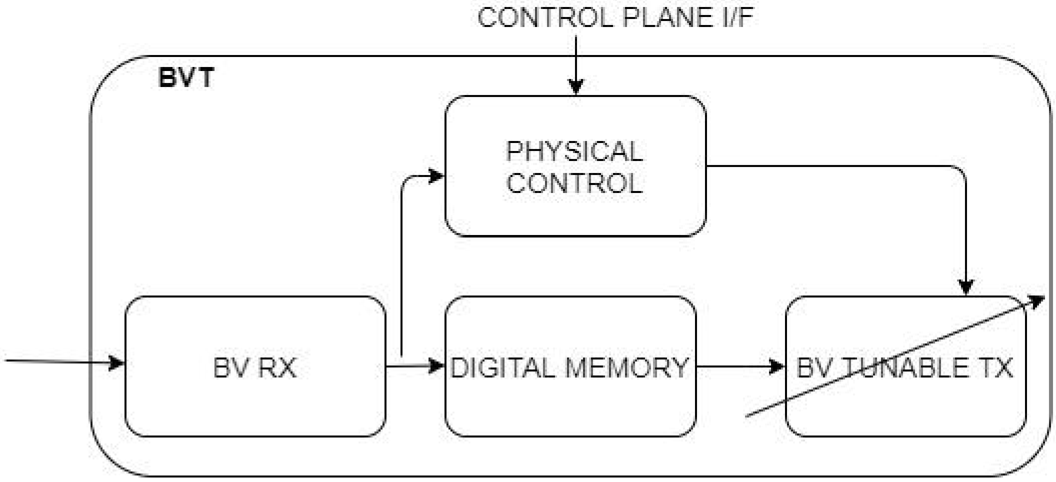


Figure 4. 5 Block diagram of a BV-Transponder.

Orthogonality of subcarriers is maintained by the principle of keeping the baud rate equivalent to the subcarrier spacing, so that a single subcarrier can have a transmission rate (*TR*) which can be represented by the subcarrier frequency slot or bandwidth (*B*) and the modulation order (*M*), which can be expressed as equation 4.3.

$$TR_{subcarrier}\,(Gbps) = B * log_2 M \tag{4.3}$$

OFDM minimum bandwidth slot of a subcarrier is selected to be 12.5 GHz, furthermore, the transmission rates are defined as 12.5 Gbps, 25 Gbps, 37.5 Gbps, 50 Gbps, 62.5 Gbps, and 75 Gbps which corresponds to the modulation techniques BPSK, QPSK, 8QAM, 16QAM, 32QAM and 64QAM respectively. This implies that in a light path, variable number of subcarriers can exist but use the same modulation technique and also noting that the variable number of subcarriers depend on the demanded capacity. Thus, the capacity (*C*) of an optical channel in OFDM-based optical network can be expressed as equation 4.4, where (*N*) represents the number of subcarriers in a channel.

$$C = N * \; TR_{subcarrier} \tag{4.4}$$

The energy consumption of a BV-OXC can be presumed to be similar to an energy consumption of a general fixed bandwidth OXC, thus the power consumption is depending on the nodal degree (N_d), and an additional over-head power of 150W per node, which can be expressed as equation 4.5.

$$PC_{OXC}(W) = (N_d * 85) + 150 \quad (4.5)$$

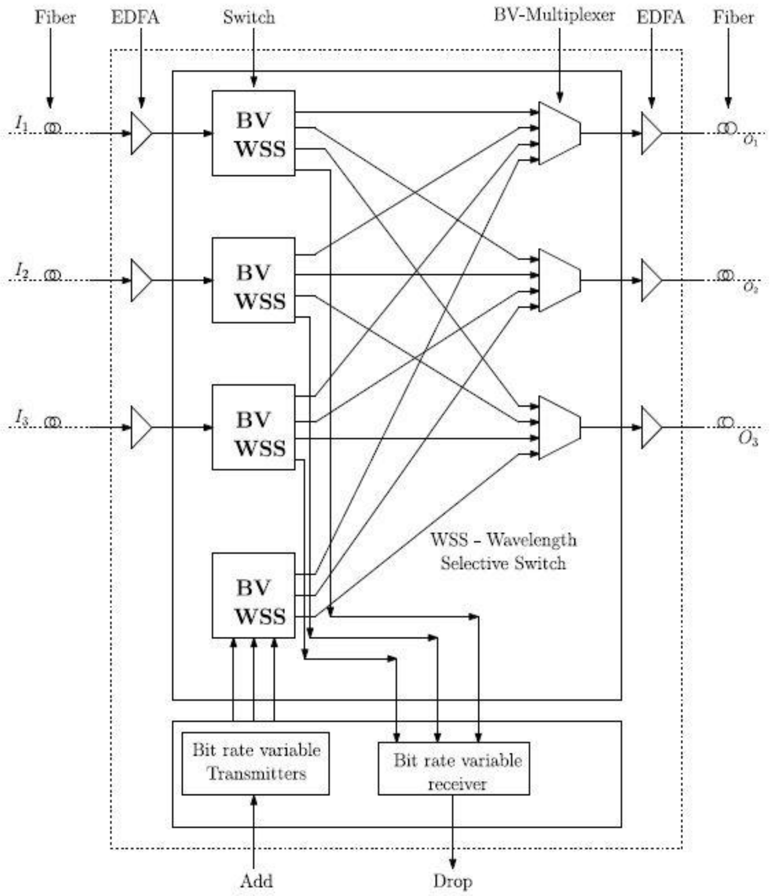


Figure 4. 6 Bandwidth variable OXC

The transponder also plays a significant role in the consumption of power for the overall optical transport network. OFDM-based optical transport network uses a special kind of bandwidth variable transponder which is referred to as Coherent-OFDM transponder. These coherent transponders are bandwidth variable and are available at both the transmitter and receiver side and they incorporate a DSP module with two digital to analog converters, which makes CO-OFDM transponders slightly more complex compared to a WDM transponder which is available only at the receiver side [78].

A single Coherent-OFDM transponder is evaluated to have a power consumption that is dependent on its transmission rate (TR) as shown in equation 4.6.

$$PC_{transponder}(W) = 1.25 * TR(Gbps) + 31.5 \quad (4.6)$$

Note that equation 4.6 is valid with some assumptions mentioned in [78]. Additionally, using equation 4.6 the energy consumption of different number of subcarriers, with their respective modulation formats can be worked out as illustrated in Table 4.2 and Figure 4.7.

Table 4. 2 Energy consumption of a Coherent-OFDM Transponder

Modulation Format	Subcarrier capacity (Gb/s)	Power consumption (W)	Transmission Reach (m)
BPSK	12.5	47.13	4000
QPSK	25	62.75	2000
8QAM	37.5	78.38	1000
16QAM	50	94	500
32QAM	62.5	109.63	250
64QAM	75	125.23	125

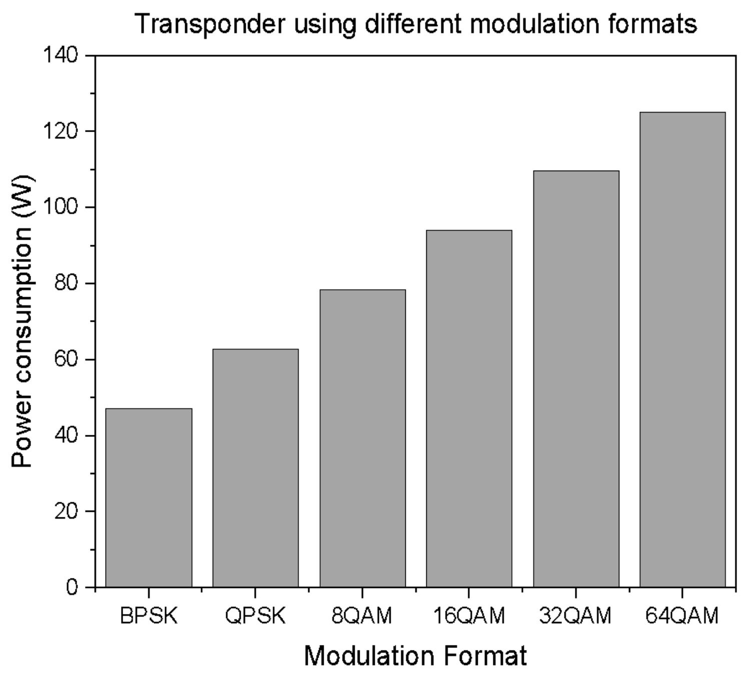


Figure 4. 7 Energy consumption of a Coherent-OFDM Transponder

In Figure 4.7 above, it can be clearly observed that the energy consumption of a Coherent-OFDM transponder increases with the subcarrier capacity. The modulation technique used depends on the desired transmission reach and the subcarrier capacity. Once these two degrees have been identified, the signal transmitted over the optical path is routed through BV-OXCs and BVTs towards the receiver end, significantly every BV-OXC on the route assigns and sets the appropriately sized spectrum cross-connection which permits a suitable end-to-end optical transmission. To achieve this flexibility, the BV-OXC has to adjust its spectral switching window contiguously and accordingly to cater for the spectral width to suit the subsequent optical signal [79]. Figure 4.8 shows a detailed elastic optical path network with the appropriately sized traffic. In this network, to achieve efficient transmission, the amount of traversing traffic is firstly determined and verified, then the network reconfigures its BV-OXCs and BVTs to match the incoming traffic load for the appropriate bandwidth and spectrum allocations. Each of the optical path can be automatically switched, and when the traffic demands vary at any given time, then the bandwidth/spectrum can be re-adjusted accordingly using BV-OXC and BV Coherent-OFDM transponders (BVT) [80].

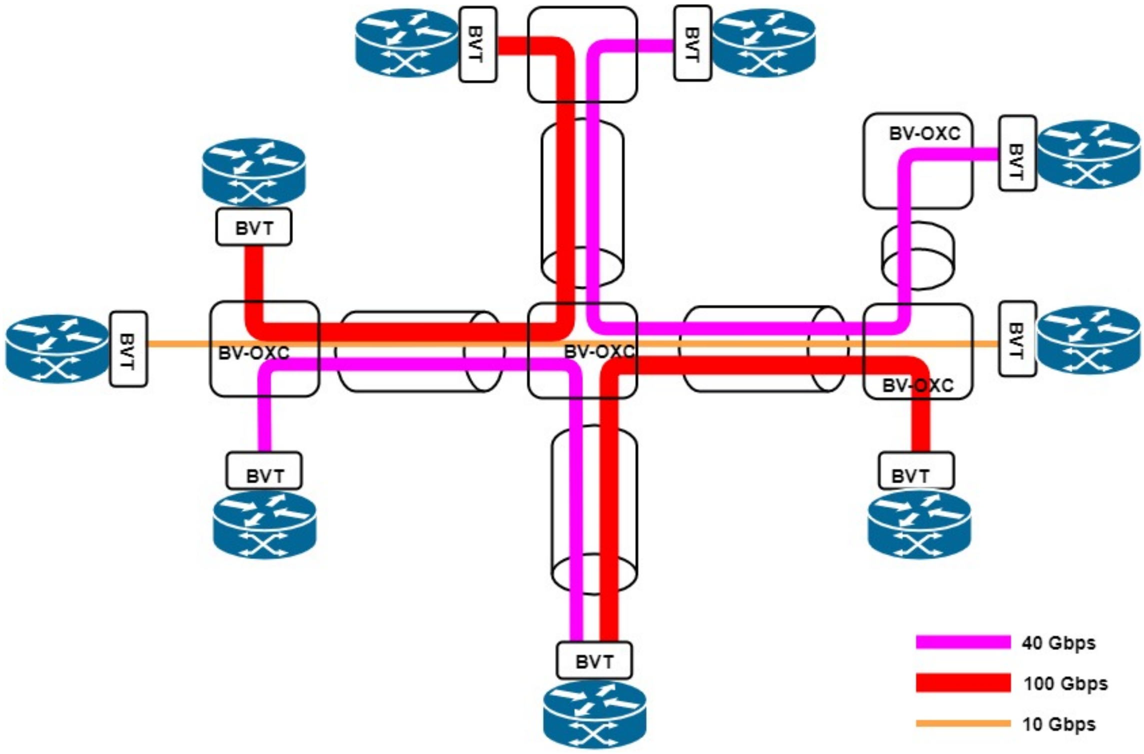


Figure 4. 8 Elastic optical path network.

4.6 Algorithms in OTNs

4.6.1 Routing and Wavelength Assignment (RWA) in WDM networks

Spectrum allocation constrains, inconsistent modulation techniques, and an increased number of subcarriers are the main differences between conventional WDM networks and elastic OFDM networks. Due to the clear differences mentioned above, this has resulted in new algorithms being adopted by elastic optical networks. Routing and Wavelength Assignment (RWA) algorithms that are used in traditional WDM optical networks are no longer suitable for an OFDM-based flexible optical network, hence a new algorithm which is applicable for elastic OFDM-based optical network is referred to as routing, modulation-level and spectrum assignment (RMLSA) algorithm.

4.6.2 RMLSA in Elastic Optical Networks.

A technique of serving requested demands one-by-one is used in this algorithm, whereby the highest priority demand is served first. Several candidate paths (k) that exist between the source node and the destination node are determined, and on each path, the potential modulation techniques are established by means of comparing the path distance with the optical reach of each modulation format, and also examining the links on the availability of contiguous subcarriers along the path. This technique allows for the appropriate combinations of the path length and modulation technique to be chosen based on the summation of end-to-end energy consumption (which includes in-line network components that are traversed in the path such as transponders, cross connects, EDFAs etc.). Then the most efficient path in terms of power and modulation technique is selected, and the corresponding adjacent subcarriers are allocated to the path based on the demanded capacity.

4.7 Advantages of Elastic Optical Network

Elastic optical networks promise to overcome many inefficiencies on today's networks to achieve a resilient, flexible and adapting network architecture for current and future networks. A summary of a few major benefits that are introduced by using elastic optical networks are presented as follows.

4.7.1 Increased Spectral Efficiency

The network capacity is increased drastically because of the optimization of the spectral allocation on demand request. Additionally, an efficient use of the deployed optical links is enabled, resulting in a much longer life span of optical links.

4.7.2 Improved Economics

Capital expenditures are reduced because of the integration of equipment and devices. Likewise, operational expenditures are also reduced because the network is more energy efficient in terms of its operation. Transponders and OXCs are bandwidth variable which means the same transponder and OXC can be used for different traffic demands.

4.7.3 Accommodation of 1 Tbps Demand

Future demands of subscriber signal transmission rates of 1 Tbps and above can never be accommodated by fixed-grid channels of 50 GHz. However, elastic grid optical transport networks can accommodate such huge capacity demands.

4.7.4 Dynamic Reconfiguration

The use of BV-Transponders and BV-OXCs on elastic optical networks enables the network to dynamically reconfigure if necessary, which is a huge advantage in current and future bandwidth-hungry network operations.

4.8 Summary Chapter Conclusion

In this chapter we discussed how elastic optical network can significantly reduce energy consumption and improve spectral optimization and consequently improving the optical transport networks to become flexible and reconfigurable. This flexibility can play a significant role in the design and operation of current and future networks such as 5G network fronthaul, backhaul and backbone core network.

5. Efficient Regenerator Placement in OTNs

5.1 Introduction

The signal strength of an optical signal or any signal in general degrades as it traverses through a light path, and this optical degradation is due to transmission impairments, which means a power boost of the optical signal is required along the path after a specific distance referred to as optical transmission reach. As mentioned in Chapter 2, in translucent optical networks, few nodes are capable of being regenerator nodes and are strategically placed to ensure that traffic requests reach their destination in an acceptable QoS. Impairments are introduced along the light path due to optical switching, optical multiplexing and optical amplification. Impairments can be categorized into non linear and linear impairments, common linear impairments are amplified spontaneous emission (ASE) noise, polarization mode dispersion (PMD) and chromatic dispersion, and common non-linear impairments are cross phase modulation and self-phase modulation. Therefore, due to these impairments, the process of placing regenerators is essential in an optical transport network. However, the regenerators must be placed carefully and effectively to reduce network cost. Effectively in a sense that if regenerators are placed too far apart then the signal might suffer too many impairments along the path and then it degrades to an unrecoverable signal.

One of the major tasks in translucent optical networks is the placement of expensive regeneration resources strategically, in order to maintain system requirements and acceptable QoS [82]. Furthermore, placing too many regenerators at every node (opaque network) even when it is not required can be costly because 3R regenerators are expensive devices in the optical transport network due to their ability to convert optical signals to electrical signals, and then amplifying the electrical signal and finally converting it back to optical signals. A typical 3R is represented in Figure 5.1. Additionally, placing unnecessary regenerators can also drastically increase the energy consumption of the entire network because regenerators consume a lot of power to boost the signal due to the process of optical-electrical-optical conversion. Hence, regenerator placement plays a significant role in power consumption of the overall network as well as OPEX and CAPEX. Generally, on an optical transport network a regenerator is placed at about 80 km of distance, however, this is not practical since impairments are not constant on every optical network. A more effective way of placing regenerators is based on the optical signal-to-noise ratio (OSNR) at intermediate nodes and following the approach of translucent optical networks in which regenerators are placed strategically where they are only required, in attempt to achieve a balance between service provisioning and network costs [83].

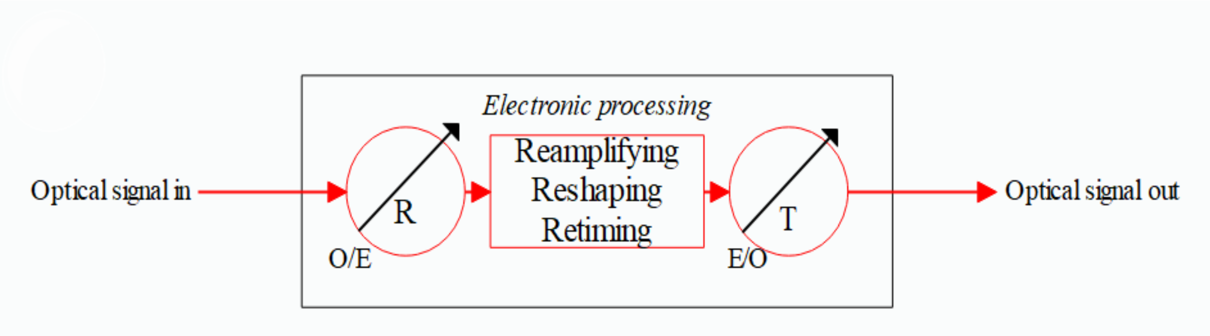


Figure 5. 1 3R regenerator

5.2 Regenerator Placement Strategies

In an optical network, a regenerator is necessary even though it is expensive. It is necessary based on the problem of optical reach as well as wavelength constrains [84], which involves wavelength conversion. Several research techniques have been discussed on literature based on the problem of regenerator placement, which is widely referred to as regenerator placement problem (RPP).

The authors at [82], proposes one way of solving the RPP problem, a technique of using an auxiliary graph approach which is based on regenerator placement and wavelength assignment (RPWA). This is effective because proper weights are assigned on the edge of auxiliary graphs, then the shortest weight path will provide the correct regenerator placement site.

The authors at [85], propose a network topology based regenerator placement, in which two algorithms are used, namely, the nodal degree first (NDF) algorithm and the center node first (CNF) regenerator placement algorithm. In NDF algorithm, the node having higher nodal degree than other nodes is assigned as a regeneration capable node and a regenerator can be placed on that node. In CNF algorithm, a node is regarded as more centered compared to other nodes if it is passed by a greater number of shortest hop distance path compared to all other nodes in a network [85], hence that node is regarded as a regeneration capable node.

The same author at [85], proposes a traffic prediction based approach of placing regenerators in which a prediction pattern is used to identify the nodes where regeneration demands are most likely required. In this way an algorithm which favors the nodes with heavier loads is used, which is referred to as traffic load prediction (TLP) algorithm. TLP based regenerator placement selects nodes with heavier traffic load as regenerator nodes. However, heavier traffic load does not mean poor signal quality, hence a more efficient strategy of regenerator placement should be based on the quality of transmission at the node as compared to node traffic.

5.3 Efficient Regenerator Placement

An efficient regenerator placement is based on the quality of transmission of the signal or rather the signal level when compared to the noise. Therefore, an optical signal to noise ratio (OSNR) based regenerator placement strategy seems to be a more logical approach towards achieving energy efficient regenerator placement.

5.3.1 Regenerator Placement based on OSNR

The OSNR is a measure of the ratio of optical signal power to the power of the system noise. OSNR is measured in decibels (dB). The OSNR is helpful because it shows the level of noise when compared to the wanted signal as the signal travels along the path. The OSNR can be expressed mathematically as equation 5.1 and equation 5.3. Figure 5.2 shows the OSNR at the source node as compared to the OSNR at the destination node after a specific distance in km has been traveled by the optical signal.

$$OSNR = \frac{P_{signal}}{P_{noise(ASE)}} \quad (5.1)$$

$$OSNR_{dB} = 10log_{10}\left(\frac{P_{signal}}{P_{noise\ (ASE)}}\right) \quad (5.2)$$

$$OSNR_{dB} = P_{signal}(dBm) - P_{noise(ASE)}(dBm) \quad (5.3)$$

Where:

$$P_{noise(ASE)}(dBm) = -58\ (dBm) + G(dB) + NF(dB) \quad (5.4)$$

P_{signal} in dBm is the signal level and $P_{noise(ASE)}$ is the noise level which is referred to as amplified spontaneous emission (ASE) noise and is dependent on the gain (G) and noise figure (NF) of the optical amplifier.

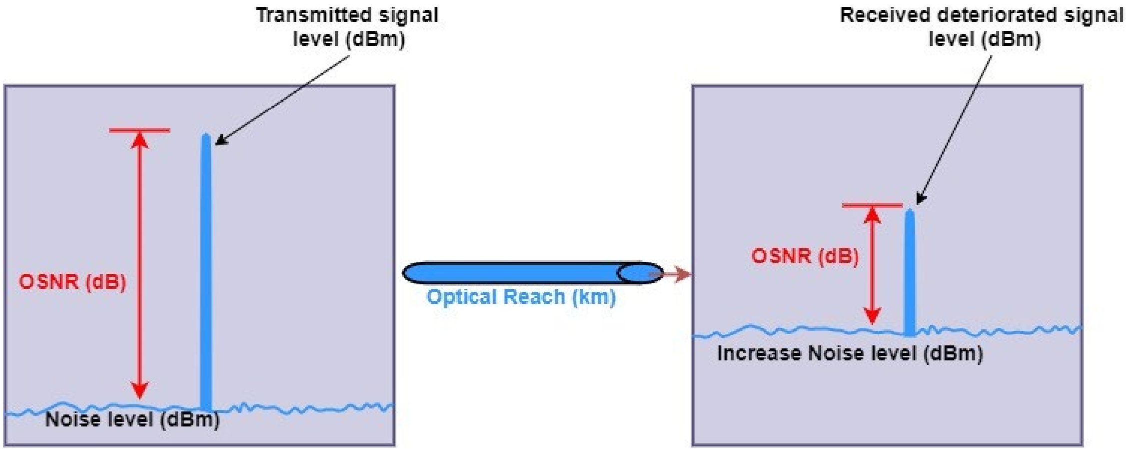


Figure 5. 2 Degradation of signal level as it travels through a light path

The process of placing regenerators effectively from source to destination can significantly minimize the consumption of energy. Therefore, the proposed technique based on OSNR will greatly minimize power consumptions in optical transport networks. As mentioned before that this approach is based on the measured OSNR, so to determine where to place a regenerator, starting from the source node, an OSNR is calculated at intermediate node, and if the OSNR at the next intermediate node is greater than a predefined threshold, then there is no need to place a regenerator at that node, then the OSNR is evaluated again at the next intermediate node. However, if the OSNR at a specific intermediate node is less than a predefined threshold, then a regenerator is placed at a node just before the evaluated one. In this sense the optical signal is regenerated only when amplification is required, and that approach helps to minimize the number of unnecessary regenerators while also maintaining an acceptable QoS. This process is repeated on the route until the destination node is reached. Figure 5.3 shows the proposed regenerator placement strategy based on the measured OSNR.

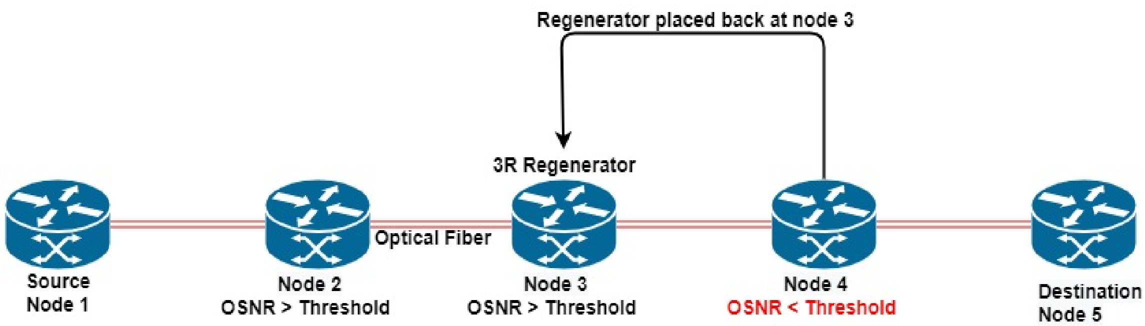


Figure 5. 3 Regenerator placement strategy on a light path

In Figure 5.3, the evaluation shows that at node 2 and 3 the OSNR is at an acceptable level above the threshold, but at node 4 the OSNR is below the threshold therefore a regenerator is placed back on node 3. Then the next starting point now is node 3, then the OSNR evaluation process is repeated until the destination node.

5.4 Summary Chapter Conclusions

In this chapter we discussed an energy efficient approach of placing regenerators in a translucent optical transport network. We proposed a technique of placing regenerators based on the OSNR, this method proves to be more efficient in terms of energy and resource allocation as compared to placing regenerators at every node. This technique offers advantages of maintaining an acceptable QoS on an optical transport network while also minimizing power consumption by using the least number of regenerators only where regenerators are required.

6. Towards Energy Efficient Design and Operation of Transport Networks.

6.1 Motivation

As more devices (mobile subscribers, multimedia applications, data applications) and terminals connect to the back bone transport network, this amongst other driving factors has necessitated design tendencies that will promote both energy efficiency operation as well as high throughput with consistent QoS. The significant driving factors that prompted towards designing an energy efficient transport (wireless and optical) network for the future are based on the growth rate of mobile internet subscribers (data traffic volume) per day which is drastically increasing every day. Globally, in 2021, the internet subscribers have been increased from 43.5% to 59.7% [87]. Most available literature on transport network focus on improving QoS and QoE but rather neglect the increased number of data traffic volume which emerged from increased mobile internet subscribers. Figure 6.1 shows the rate at which mobile subscribers have increased since 2018 towards the predicted number in 2025. The increase of mobile internet subscribers (online users) calls for an improvement in network architecture and network operation.

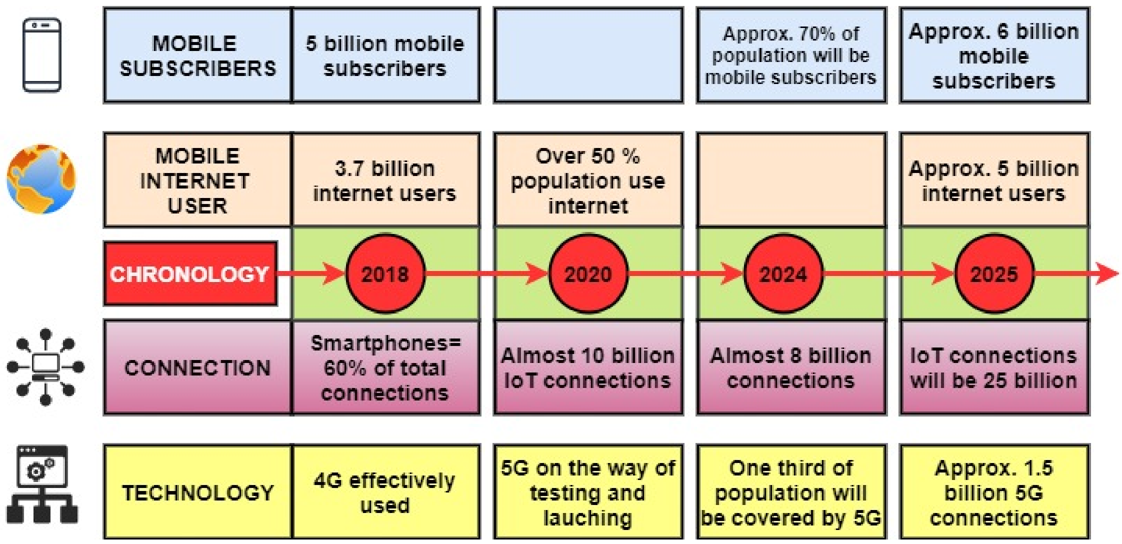


Figure 6. 1 Mobile subscriber growth of the telecommunication industry [87]

In today's network, various applications leverage on live streaming, which includes voice over IP (WhatsApp calls, Skype call or Zoom), online video applications (e.g. YouTube, ShowMax or Netflix), online conferences and online TV (e.g. DSTV now). Traffic related to video distribution is currently accountable for 66% of the total data traffic and is projected to increase even further by 2026 to approximately 77% [88].

Streaming services require faster download speed (typically over 1 Gbps) to support higher quality streaming without pauses during play black (also referred to as buffering) [86]. This rapid increase in mobile internet subscribers has resulted in an increased number of streaming subscribers, so this means it is absolutely imperative to have an efficient transport network to cope with high volume of data transmission of over 5 billion mobile subscribers with internet access by the year 2025. According to [88], Globally, the growth in mobile data traffic per smartphone is due to data-intensive content (video) and improved device capabilities. Furthermore, COVID-19 has accelerated online streaming (data traffic) due to staying indoors working remotely to fulfill personal needs and business needs, hence the average monthly usage of mobile data continues to show robust growth and will continue even higher by 2030 [88] as depicted in Figure 6.2.

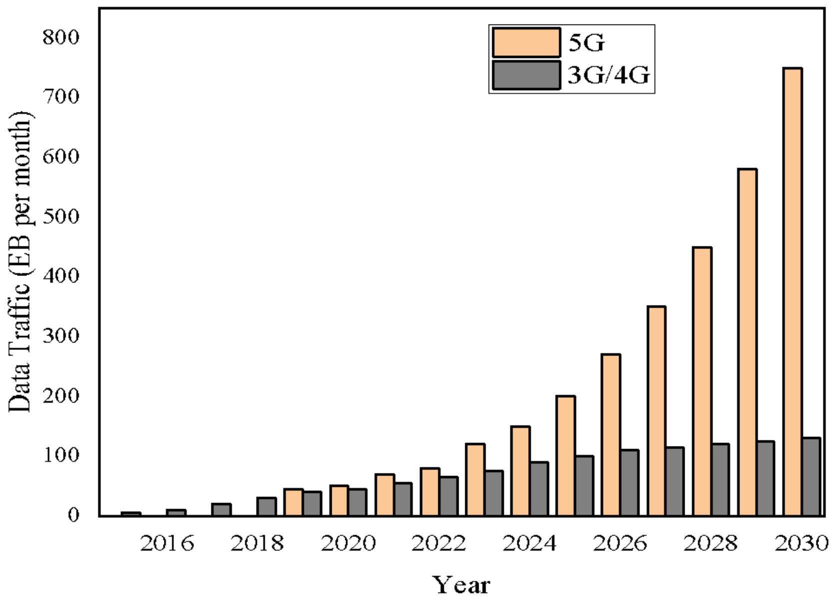


Figure 6. 2 Worldwide mobile data traffic growth (EB per month)

Moreover, according to [87], mobile subscribers are increasing rapidly every day, and in 2020 approximately 51% of the total carbon dioxide emission was due to the telecommunication system alone, and if this value is neglected, then by 2025 it would be an even higher value since mobile subscribers would have increased drastically to over 5 billion. Figure 6.3 depicts variations in CO_2 footprint produced by different sections of the telecommunication system from 2010 to the expected value in 2030. This rapid increase in CO_2 emissions has also motivated for an energy efficient transport network to curb carbon gas effects. Finally, according to [87], the electrical power consumptions by the information communication technology will be rise from 611 TWh to 1752 TWh by 2030, this extreme consumption of energy by transport networks especially wireless transport network will result in vital issues in the future if no precautions are taken to curb the existing trends. Hence, the major goals for designing an energy efficient transport network is to reduce consumption of energy of current networks, which can further reduce the operational expenditure while maintaining high throughput with consistent QoS, and finally reducing the CO_2 emission of the communication system.

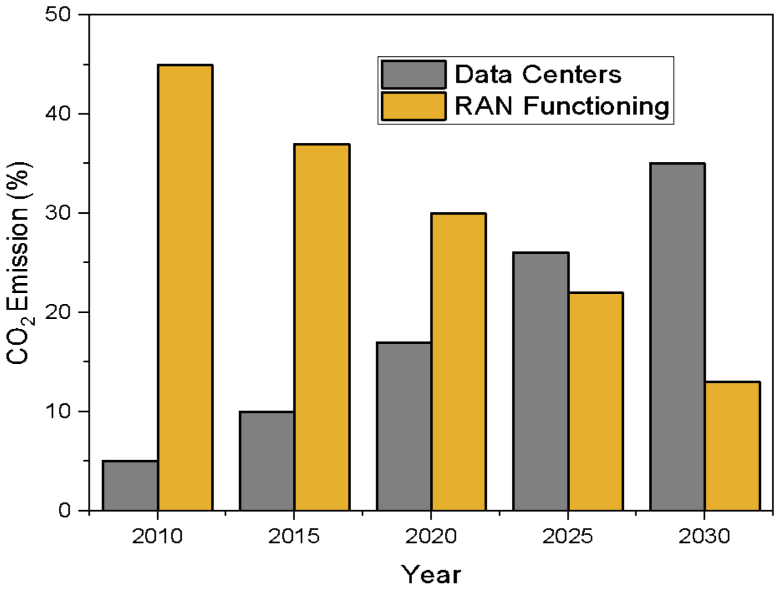


Figure 6. 3 CO_2 emission by the communication system [87]

Moreover, efficient transport network is vital in the health industry as well, currently the world is facing the problem of aging population, as well as patients with limited autonomy or chronic diseases. This increase number of patients has resulted in an imbalance of medical resource supply and demand [92]. This has motivated for the emergence of remote e-health, which could also have a huge impact on reducing the spread of COVID-19. E-health refers to the use of Information Communication Technology (ICT) to provide medical services [93]. Figure 6.4 depicts key requirements for e-health and its objectives. E-health is regarded as a critical mission service because it requires ultra-low latency, improved QoS and improved security. To achieve these high demands of e-health, an efficient backbone transport network is absolutely imperative.

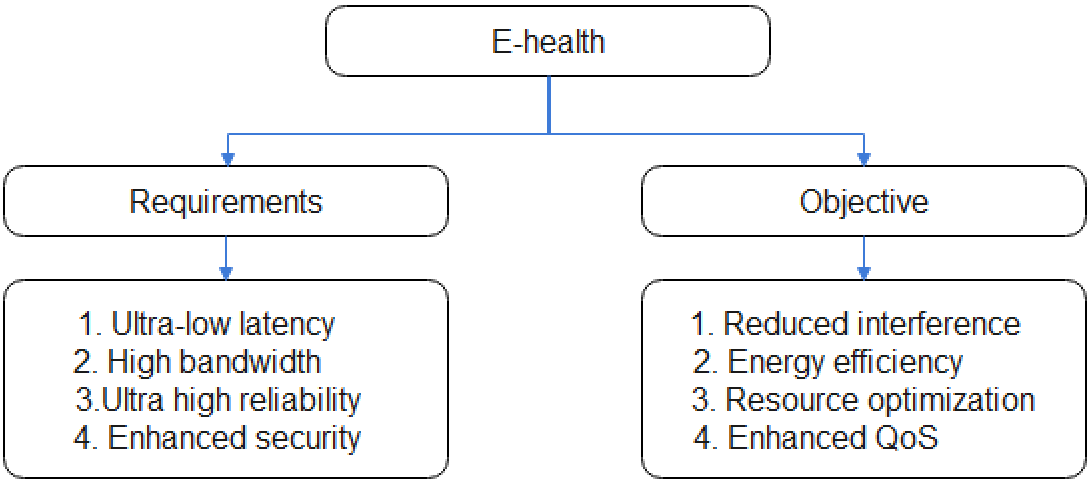


Figure 6. 4 E-health requirements and objectives

In applications such as e-health where some services are critical and require real-time transmissions then delays cannot be tolerated hence the need for low latency, improved security and high QoS backbone transport networks. For example, services like tele-monitoring (remote monitoring of patients which are diagnosed with serious chronic diseases) and tele-diagnosis (remote diagnosis of patients) require real time transmission as well improved quality therefore the backbone transport network cannot tolerate delays or packet loss. Finally, to maintain patient confidentiality, complex security systems are vital on the network when transmitting medical documents. An elastic, efficient and secure back bone transport network is thus imperative in such applications to offer robustness, enhanced QoS, ultra-low latency and enhanced security.

6.2 Desirable Design and Operation Features for Transport Networks

Modern networks such as 5G and applications such as Internet of Things (IoT) are highly sensitive to delays as depicted in Figure 6.5. Therefore, the desirable features of today's network are based on high throughput and low latency at consistent QoS, as well security. Transport Control Protocol (TCP) plays a significant role in achieving high throughput at low end to end latency. In simpler terms transport control is a network technique aimed at improving latency and throughput of the entire network [89]. TCP is widely referred to as a connection-oriented transmission which means the sender waits to receive an acknowledgement signal packet within an allocated timeout period [89], otherwise retransmission occurs. Furthermore, operation features for improving latency and throughput are based on traffic congestion, hence the need for congestion detection, congestion notification and congestion control in current and future transport networks.

A. *Congestion detection* is essential in network operation due to detecting loss of packets, traffic load, packet processing time, and throughput measurement.
B. *Congestion notification* is essential to notify transmitters about traffic load. Transmission Time Out as well Not-Acknowledgement are the main paradigms generally used for traffic notifications [88].
C. *Congestion control* is essential to predict proper paths in order to avoid congestion problems.

Moreover, in wireless networks there are many factors that influence latency and throughput, although congestion is one of them, but so as interference and signal level which also have the ability to alter network throughput and latency due to data loss. Additionally, due to mobility and channel fading, round trip time fluctuations are common in wireless networks [89]. D2D communication as well as machine-to-machine (M2M) communication are some of the applications which aim to reduce end-to-end latencies by communicating directly from one device to another, and MEC or fog computing also reduces latency by moving servers in close proximity to the users.

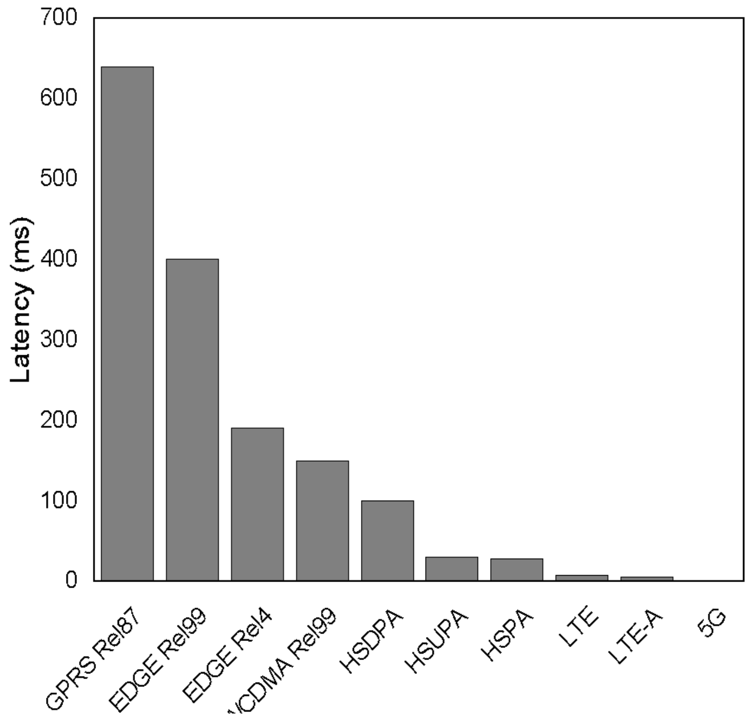


Figure 6. 5 The evolution of latency throughout cellular generations [94]

As aforementioned, routing protocols play a significant role in improving latency and throughput, hence efficient routing protocols are absolutely imperative in today's network. Routing protocols are categorized into reactive and proactive protocols. Reactive routing protocol is also referred to as "on-demand" routing protocols whereby a route is searched only if a node demands to send data to any other particular node [90], however, proactive routing protocols are table driven, meaning all the nodes maintain all the routing information for all other nodes in a routing table [90]. There are various routing protocols available on literature, the author in [90] discusses Dynamic Source Routing (DSR) protocol which can be adopted in wireless networks, Ad hoc On-demand Distance Vector (AODV) protocol, as well as Destination Sequence Distance Vector (DSDV) protocol. Finally, a promising approach for modern networks is a technique known as SDN-based routing protocol. Since SDN provides an entire view of the underlying network, then transmission delays can be avoided while maximizing packet delivery which results in high throughput and low latency. The author at [91] proposes an SDN-based connectivity and geographical-aware protocol for routing in which the SDN controller computes routing evaluations in real time using mobile networks.

6.3 Architectural Design

In modern networks such as 5G there are numerous architectural requirements to support mission critical communication. Reliability and delay optimization are two most crucial amongst them. Today's network architecture must achieve flexibility in the transport and control domain, this is made possible by the use of emerging technologies such as SDN and NFV. Furthermore, current network architectures must be capable to produce high throughput for services such as online streaming, and this is made possible by the use of small cells to provide ubiquitous coverage. Additionally, the network architecture should be energy efficient and this is achieved by the adoption of D2D communication and massive MIMO antennas on base stations, and finally the wired network such as front-haul and backhaul must also be flexible hence the use of OFDM based optical network as a front-haul, backhaul and backbone network. Today's network architecture should also take advantage of cloud services and edge computing to achieve ultra-low end to end latency. Today's architecture must be elastic and take advantage of all the aforementioned requirements for modern networks, and hence making it an efficient architecture to cater for current and future service demands on a network. These elastic architectures are based on Cloud RAN, edge computing or fog computing, flexible OFDM front-haul, back-haul and backbone network, and an SDN and NFV based optical core network. An example of an elastic architecture for modern networks is represented in Figure 6.6, in which the building block are discussed as follows.

- *Cloud RAN*: This is an emerging paradigm which leverages on the benefits and characteristics of cloud in the network which results in low latency, improved energy efficiency, improved network capacity, and optimal resource management [94].
- *SDN/NFV backbone core*: The backbone core network functions such as mobility management function (MMF), authentication functions (AF), session management functions (SMF) and policy control functions (PCF) are migrated to the SDN network controller as applications (virtual functions) rather than actual hardware components. The forwarding switches are referred to as SDN switches since they communicate directly to the SDN controller via open-flow protocol.
- *BBU pool*: This is a technique used to migrate baseband processing to a centralized location known as the BBU pool which results in better processing and management.
- *Remote radio head (RRH):* These form part of the radio access network, and can be regarded as low power base stations with the baseband processing stripped away from them. RRH are used for small cell coverage to improve data throughput, and these are used together with conventional macro base station to form a heterogeneous network with ubiquitous coverage. Furthermore, RRH support massive MIMO antennas to further improve the network efficiency.
- *Front-haul*: This is the interconnection between the RRH and the BBU pool, it can either be wired or wireless network. This is a delay in-tolerant interface because it transports both data and control information. In this architecture a wired approach using flexible OFDM-based optical interface is adopted in the front-haul to provide optimal resource allocation as shown in Figure 6.7 as compared to conventional fixed common public radio interface (CPRI).
- *Back-haul*: This is the interconnection between the backbone network and the BBU pool. This is also designed based on flexible optical interface.

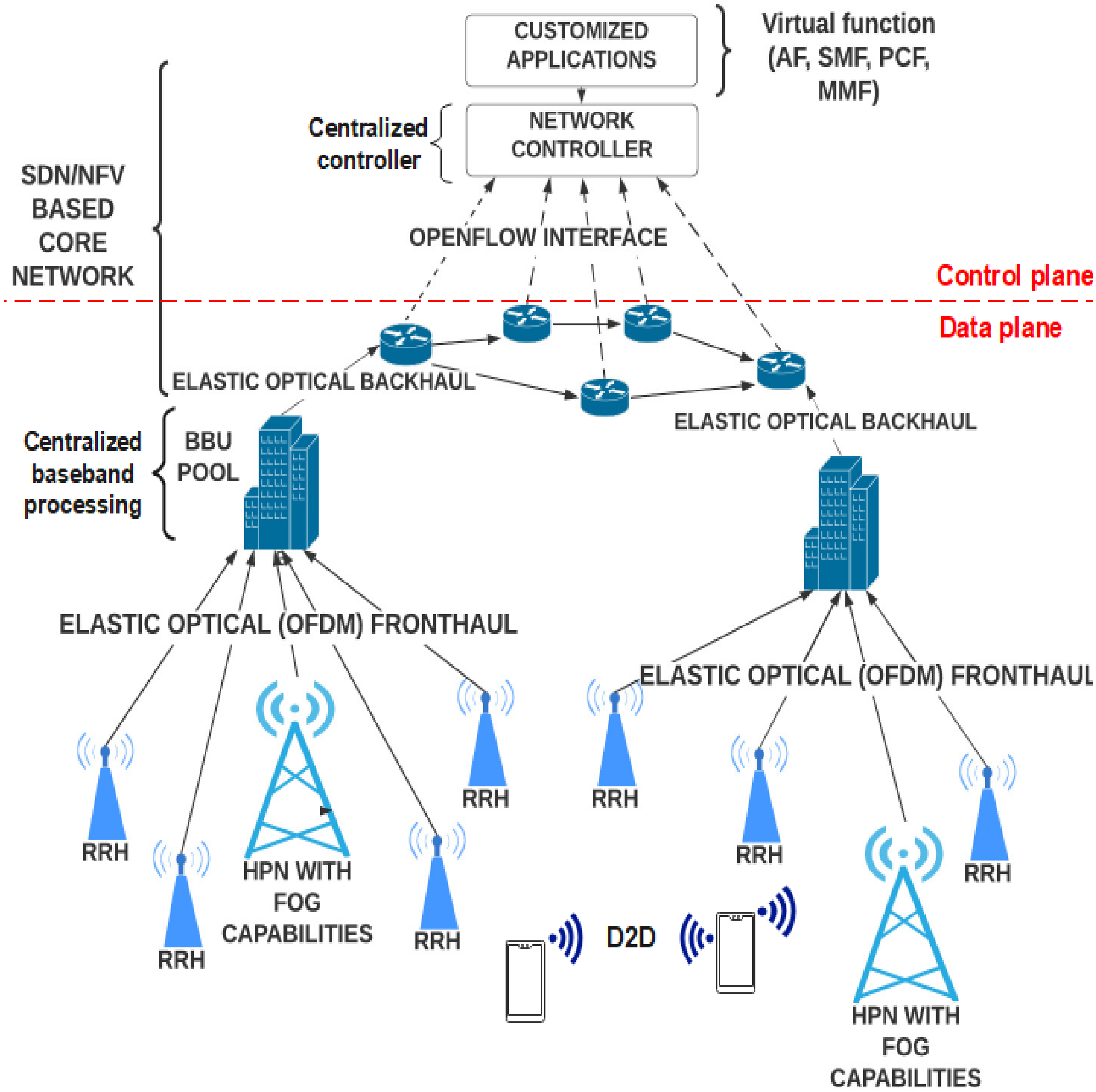


Figure 6. 6 Elastic architecture

Moreover, the presented elastic architecture uses RRH and high power nodes (HPN) with fog computing or caching capabilities as well as resource management functionalities. This means that the caching capability on the RRH relieves the constant transmission of data through the front-haul and results in low latency. The use of smart devices which support D2D communication also reduce latency which is why it is included in the elastic architecture. These smart devices also have caching capabilities as well as radio resource management functionalities which allow these devices to communicate directly with one another or simply relay information which also results in a relieved front-haul.

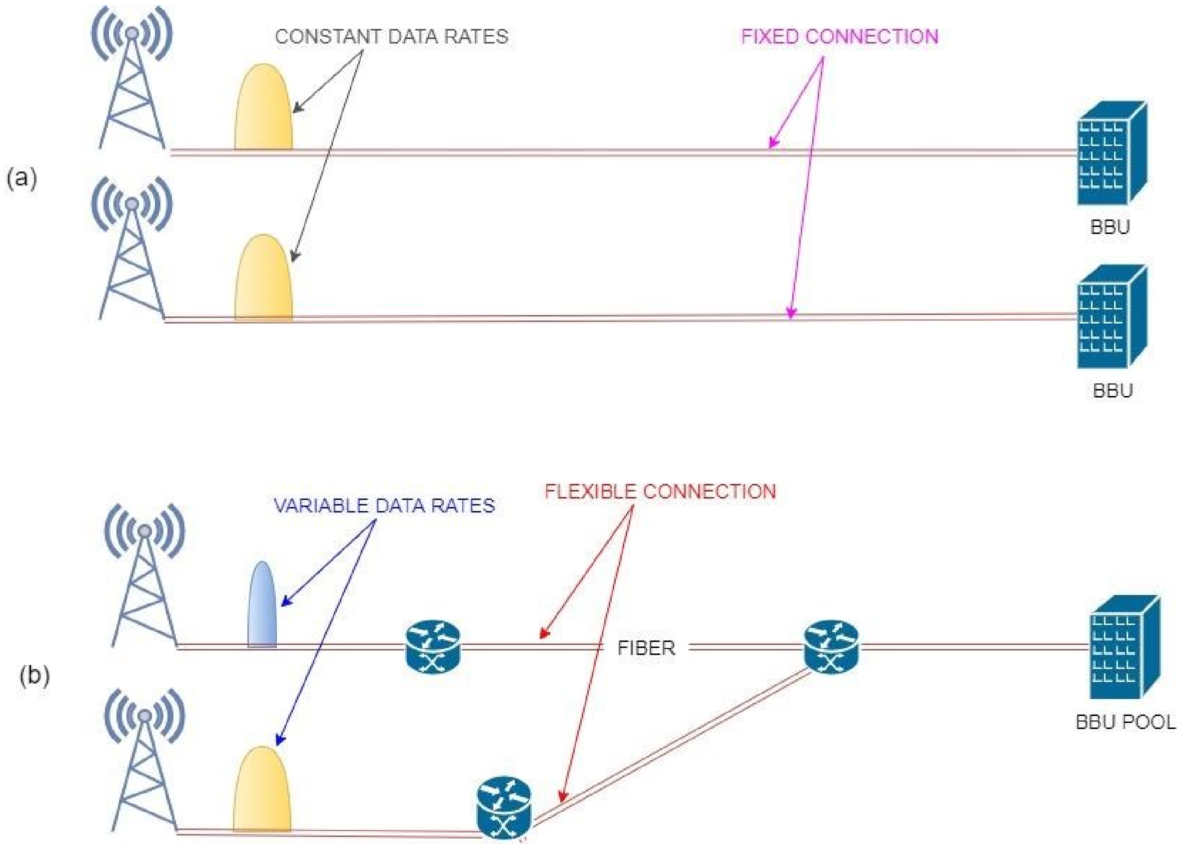


Figure 6. 7 (a) Conventional Front-haul vs (b) Flexible OFDM-based Front-haul

6.4 A Joint Photonic/Wireless Transport Network Architecture

The proposed network architecture required to support high-speed, ultra-reliable and low-latency data links for 5G and beyond is referred to as a joint photonic and wireless transport network, which comprise of an all-photonic network together with a wireless network as depicted in Figure 6.8. This is an advanced network approach that can efficiently process large amount of data, and this proposed architecture enables the use of ultra-high speed optical transmission and wireless transmission by connecting building (factories, hospitals, etc.) and data centers using wideband optical networks and wireless access network in an end-to-end manner. The proposed joint all-photonic and wireless transport network architecture introduces photonic-based technology to everything from networks to terminals, which results in large capacity, low delay optical multicast path. Furthermore, the proposed network architecture provides a closed feedback path that supports mission-critical services such as remote monitoring and control services for autonomous vehicles and e-health.

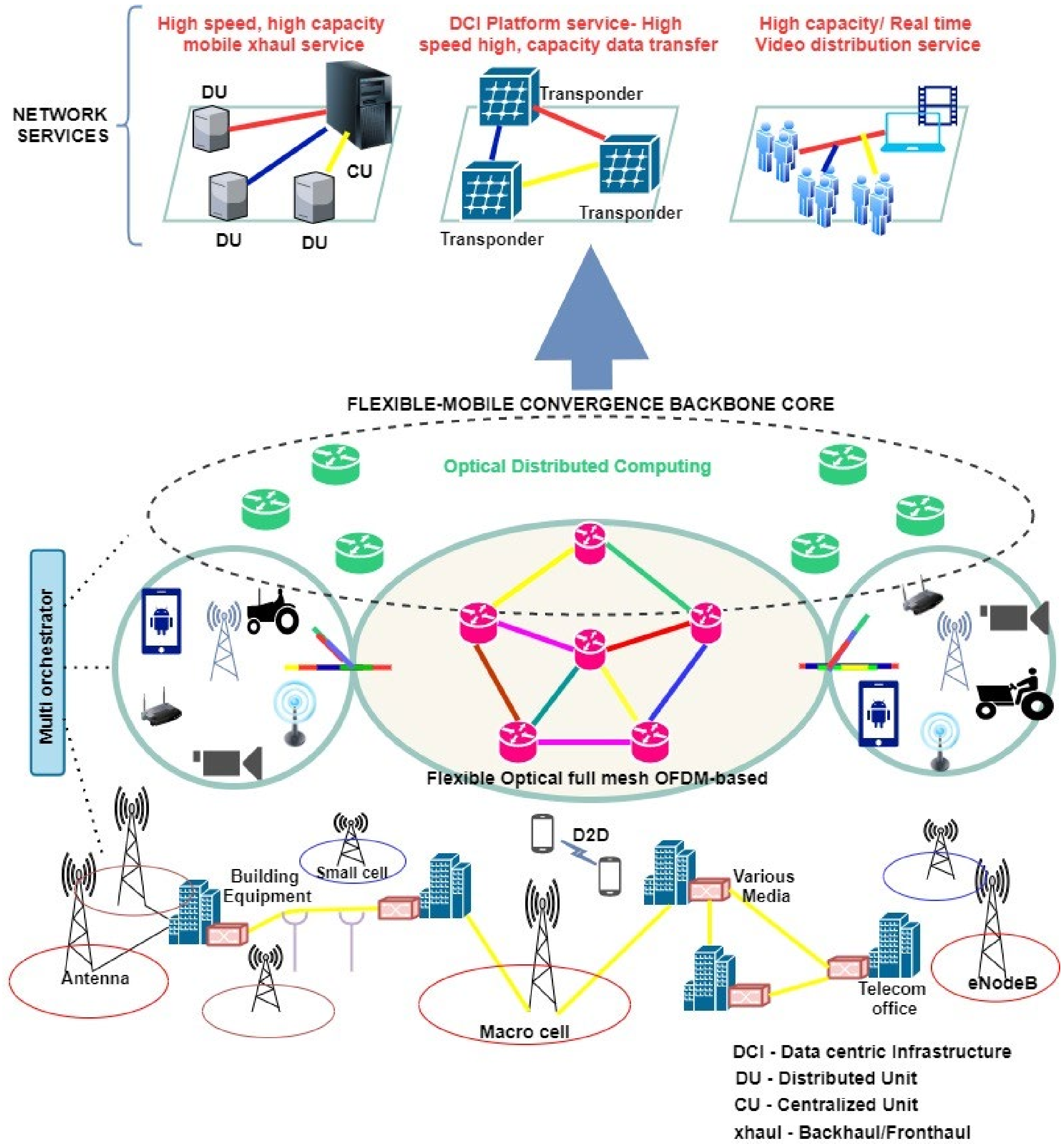


Figure 6. 8 A joint Photonic and wireless transport network architecture

The multi-orchestrator is the architecture's central manager that controls several resources such as controlling the data-centric infrastructure, controlling the network functions, controlling the wireless-infrastructure and also controlling the photonic-infrastructure. Furthermore, it is also responsible for resource management, failure monitoring, transport resource switching and analysis.

6.5 Operational Modes for Energy Efficiency

The network operation can either be centrally operated or distributed, however in a joint all-photonic and wireless transport network architecture a centralized approach is opted because of using a centralized controller (multi-orchestrator) to achieve an energy-aware network operation. The centralized controller is significant to retrieve network knowledge in order to run algorithms based on retrieved traffic loads, power state of the device and also estimate future traffic. The main criterion for energy-aware network operation is based on shutting down targeted and unused devices and also leveraging on the dynamics of routing in the optical layer.

6.5.1 Energy Efficient Operation for Wireless Transport Networks

In order to achieve energy-aware network operation in wireless transport networks, targeted wireless resources such as base stations (RRH and HPN) are turned off during idle mode, this is because traffic load conditions change with time and space in a specific region, therefore, underutilized wireless resources can be automatically turned off [87], however the user can still be covered by active wireless resources to maintain QoS. The proposed base station sleep mode operation is illustrated by Figure 6.9, noting that a base station in this case can be any wireless node (RRH/access point/ HPN). It is clear that the sleep mode is dependent on the traffic load at each base station but also noting the significance of maintaining an acceptable level of QoS.

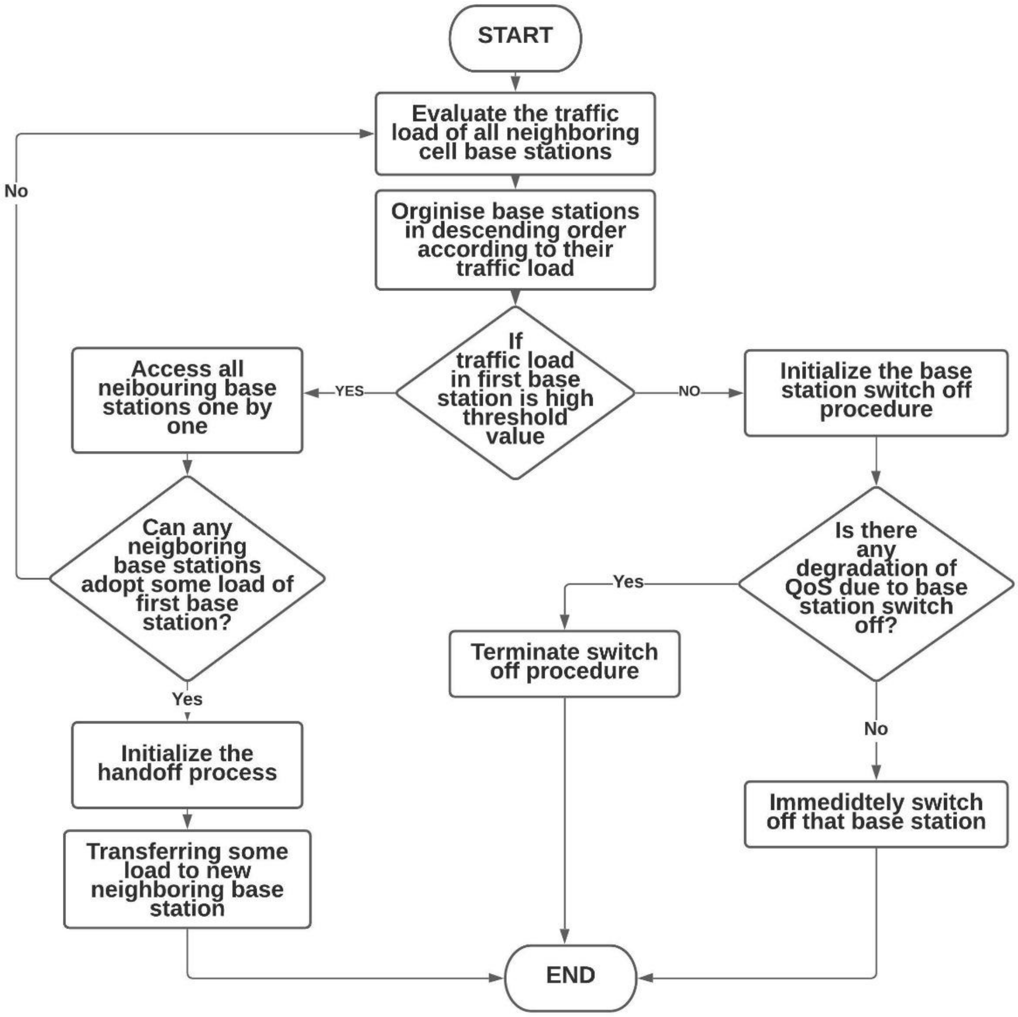


Figure 6. 9 Sleep mode operation

The other approach used in the proposed all-photonic and wireless architecture to achieving power efficiency in wireless transport networks is the adoption of D2D communication. This technique greatly reduces power consumption because D2D communication bypasses the use of cell towers (or base stations) to support proximity services, therefore resulting in high power consumers (i.e. base stations) being bypassed to allow direct communication between two or more devices.

6.5.2 Energy Efficient Operation for Optical (Photonic) Transport Networks

On optical transport networks, the efficiency in operation is achieved through the dynamics of efficient routing, thus efficient routing algorithms are imperative in modern all-photonics networks. In the architecture the front-haul and backhaul uses flexible optical transport network which is based on OFDM. This technique provides advantages on spectrum and resources allocation which results in efficient optical transport networks. Physical layer impairments-aware routing and spectrum assignment algorithm (PLI-aware RSA) is used to achieve efficient network operation on elastic optical transport networks. PLI-aware RSA algorithm considers three aspects before choosing a light-path, namely, routing, spectrum allocation and quality factor. Therefore, the path is chosen based on the highest Q-factor and shortest routing path and then the appropriately sized frequency slot is allocated.

A. Routing in PLI-aware RSA

The n-number of paths (or routes) from source (*s*) to destination (*d*) are found, and using Dijkstra's algorithm, all the located shortest paths between source and destination are established and stored in the candidate vector according to their path length. The Dijkstra's algorithm is repeated for n-times and in the first computation of the algorithm, the first and most shortest path is located (if it exists) denoted as P_{1sd}. The algorithm computes for the second time to find the second shortest path (P_{2sd}) and this process is repeated until no more paths exist. If no paths exist, then the connection is blocked [95]. It should be noted that $P_L(P_{1sd}) \leq P_L(P_{2sd}) \leq P_L(P_{3sd}) \leq \cdots P_L(P_{isd})$, where $P_L(P_{isd})$ is the path length of path P_{isd}.

B. Frequency slot allocation in PLI-aware RSA

The spectrum of each link is represented by a series of frequency positions referred to as slots ($f_1, f_2, f_3, \ldots f_{|N|}$), where |N| represents the number of frequency slots in a fiber channel [95]. Leveraging the ability of OFDM allows frequency slots (subcarriers) to overlap, therefore allowing more compression on each fiber link. The minimum frequency slot is chosen as 12.5 GHz. To determine whether a frequency slot is available or not, a binary representation state is used for each frequency slot, for instance in this case a binary state "1" represents an available frequency slot and binary state "0" represents an unavailable (or occupied) frequency slot. A network generally consists of |A| nodes, |E| links between those nodes, and in each link, |N| slots are available. Hence, an |E| x |N| matrix is formed whereby in the beginning all vectors inside the matrix are filled with binary "1" which denotes all frequency slots are available. Furthermore, each traffic demand (T_{sd}) per connection is calculated and converted to the number of required frequency slots (η_{sd}) using equation 6.1, therefore the frequency slots are allocated to a link based on traffic demands.

$$\eta_{sd} = \frac{T_{sd}}{\delta} \tag{6.1}$$

Where η_{sd} represents the total frequency slots, T_{sd} is the traffic demand, and δ is the width of the frequency slot. If the is atleast one frequency slot available and also the spectrum constraints (contiguity, continuity, and conflict) are satisfied, only then can the connection be made.

Spectrum continuous constraints (C1): The allocated frequency slot (in GHz) must be equal for each fiber.

Spectrum contiguity constraints (C2): Spectrum slot should always be consecutive for a demanded traffic.

Spectrum conflict constraints (C3): Exact bandwidth slots cannot be used on the same physical link.

Bandwidth slot allocation can be achieved using various approaches such as longest path first (LPF) allocation described in [97] and first-fit (FF) allocation describe in [96].

C. Q-factor measurement in PLI-aware RWA

Once the allocation of bandwidth slots is accomplished, then the quality factor at the destination (receiver) side is measured on all the n-paths established at the routing stage. The selected path must have the highest Q-factor compared to all n-paths with $Q \geq Q_{th}$, where Q_{th} is the threshold quality factor. If all paths don't satisfy the condition $Q \geq Q_{th}$ then no connection is made. This process enforces the QoT before any connection is established. This means a flexible path (flex-path) is only established when the quality factor is greater than the quality threshold [95]. The PLI-aware RWA algorithm is described in algorithm 2.

Algorithm 2 PLI-aware RWA algorithm

Input	:	G= (V,E) Frequency allocation matrix (U), Connection request CR((s,d),nsd,ta,th),k,Qth.
Output	:	Flexpath between (s,d), satisfying the constraints C1, C2, and C3 with Q ≥ Qth.
1	:	**for** each connection request **do**
		{
2	:	Compute the k-shortest paths by using Dijkstra's algorithm from s to d.
3	:	**If** no route exist, go to Step 13
4	:	Calculate the unused frequency-slots in every link which are present along the shortest-path found in 2.
5	:	Repeat Step 4 for all k-candidate paths.
6	:	Find a path that has nsd number of frequency-slots and satisfies the constraints C1 to C3. If no such path exists, go to Step 13.
7	:	Allocate the frequency-slots for all the candidate paths by using FF, LPF, MSF, and PFSD schemes
8	:	Measure the Q-factor for each candidate path.
9	:	**If** no path satisfies the condition Q ≥ Qth, go to Step 13.
10	:	Select the path which has the highest Q-value and Q ≥ Qth.
11	:	Establish the connection by using the path selected in Step 10.
		}
12	:	end **for**
13	:	Block the connection request

6.6 Analytical Model for an All-Optical backbone Network

The switching protocol opted in this optical core backbone network is referred to as **optical burst switching (OBS)**. OBS leverages characteristics from both wavelengths routed networks and optical packet switching networks to establish a holistic burst switching optical network. The fundamental switching technique in OBS is based on a burst. A burst is essentially a series of packets under the same header traversing and switched together from a source node to a destination node via intermediate nodes. This burst comprises of two entities, the header which is normally referred to as the control burst (CB) and the data burst (DB). The Control burst is obviously transferred first because it is responsible for bandwidth reservation along the optical path to create a room for transmitting the corresponding DB afterwards. The two transmissions between the CB and the DB are separated by a time delay which is referred to as a burst offset time. It should be noted that the DB is transmitted through the same path, which was booked by the CB, which then results in no buffering required for the DB at central nodes.

The are two main signaling protocols for OBS which can be used, i.e. Just-In-Time (JIT) protocol and the Just-Enough-Time (JET) protocol. The JIT protocol functions based on the arrival of the CB on the core node, and once the CB has arrived on the node, immediately a wavelength channel is reserved but only if it is available, otherwise if all wavelength channels are occupied hence no channel is available then the request in denied, consequently the corresponding data burst is regarded null and void. However, If the wavelength was available hence the reservation was successful then the reserved wavelength will remain reserved until the corresponding DB has been fully transmitted, hence network nodes must be able to retain information whether the wavelength reservation is still required or not. The JET protocol functions based on a fixed time span approach that is reserved for transmission once the CB has arrived. In this protocol the CB contains the data burst size (*b*), destination address, burst offset time (*a*), burst arrival time, and the wavelength whereby the corresponding DB will arrive at. The wavelength channel scheduling (WCS) algorithm is triggered once the CB has arrived at the core node, this WCS algorithm is responsible for locating the appropriate wavelength channel along the path for the incoming corresponding data burst. Then the reserved wavelength channel is only booked for a time span equal to the burst size, this time duration starts immediately once the DB has arrived. The scheduler receives information from the CB to keep updates on the availability of unused time intervals on every wavelength channel. The scheduler decides whether the wavelength channel is unscheduled based on the time (*t*) if no burst is occupying that channel at that time (*t*) or after, consequently that channel is said to be unused, which can be represented by Figure 6.10.

6.6.1 Blocking Performance

In OBS core backbone networks, the probability that an optical wavelength reservation request will be denied due to the unavailability of unused optical wavelength channels is referred to as the blocking probability, which is the essential performance measurement in OBS networks. The key factor in determining an acceptable blocking probability of the OBS network is based on the signaling protocols i.e., JIT and JET. Using the Erlang B-formula for the loss probability, then the blocking probability for an OBS-JIT network can be defined as follows:

$$B_p\,(p,k) = \frac{p^k/_{k!}}{\sum_{i=0}^{k} p^i/i!}$$

$$B_p\,(p,k) = \frac{[\curlywedge(a+b)]^k/_{k!}}{\sum_{i=0}^{k}[\curlywedge(a+b)]^i/i!} \qquad (6.2)$$

Where k represents the number of wavelengths, and p represents the load intensity. The load intensity can be further represented by $\lambda\,(a+b)$, based on the analysis that each request jams the channel for a period equal to the combination of the burst transmission delay and the burst offset time, where λ represents the mean arrival rate, a represents the burst offset time, and b represents the data burst duration.

Erlang B-formula can also be used to estimate the blocking probability of the OBS-JET network, but the issue here is that the offered load intensity will be represented by $\lambda\, b$, meaning that the effect of the offset size is neglected. Therefore, the Erlang's equation is not a good approximation for the blocking probability of the OBS-JET network. A good approximation should include all variables for calculating the blocking probability, so for an OBS-JET network, if a link with M data wavelength channels is considered, and each CB is arriving to request a channel to be reserved. A control burst requests a wavelength channel after an offset time (a) and data burst time length (b). Assuming t_r to be arrival time for the CB, then the request is only accepted if the reservation duration ($t_r + a + b$) does not intersect with an already booked interval, otherwise if there is any overlap of intervals then the request is denied and the corresponding DB is regarded null and void. To improve the traceability of this model, the time interval is divided into equal small slots (β), and these small slots add up to form the offset time (a) as well as the data burst length (b). Furthermore, these slots are numbered with reference to the CB based on the arrival time and n is used to represent the slot number. The simplified model for OBS-JET network is illustrated below in Figure 6.10.

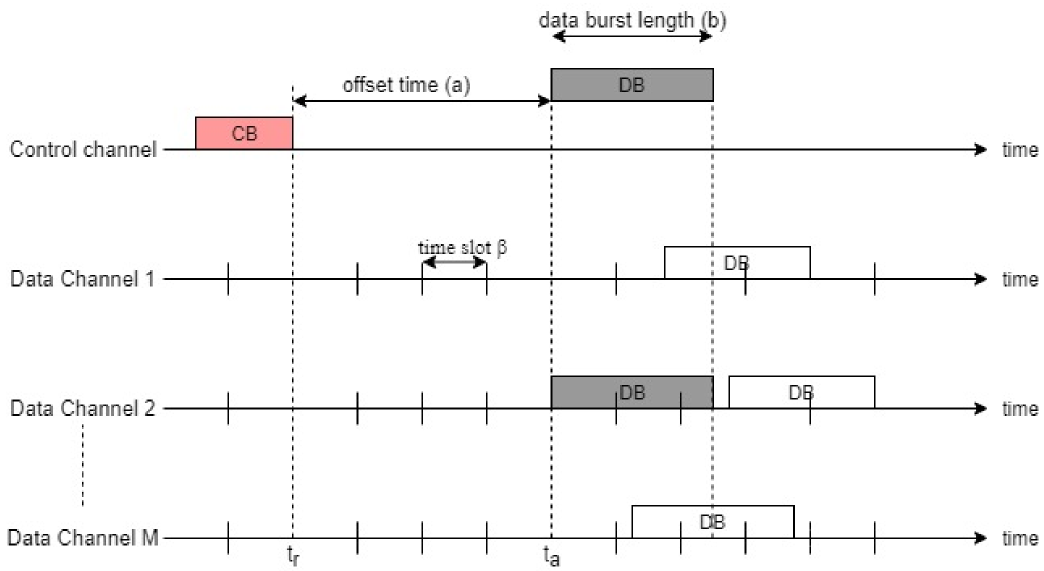


Figure 6. 10 Model for OBS-JET networks

The traffic model is evaluated based on the Poisson distribution, this is because the CB arrival process can be presumed to initiate a Poisson process with mean arrival rate (λ) per time slot. Therefore, the Poisson distributed mean rate can be expressed as follows:

$$\lambda_n = \begin{cases} \lambda, \; n \leq 0 \\ \lambda \; [1 - \sum_{a=0}^{n-1} f(a)], \; n > 0 \end{cases} \tag{6.3}$$

Where $f(a)$ represents the probability distribution of the offset time $P(a)$ for a = 0, 1, 2, 3, …. Equation 6.3 is justified based on a fact that a CB arrives a slot 0, then the probability that the arrived control burst will request (λ) a time slot *n* is $P(n)$, meaning that the total number of requests from slot 0 to slot *n* can be represented by $\lambda \; P(n)$. Basically, this means the number of arrivals in any slot *n* for a reference control burst is the summation of all potential requests (λ) subtracting any future request to be made between time slot 1 to n.

To determine the state of the wavelength channel (whether occupied or not), a Markov chain is used in which the data burst duration is assumed to have a geometric distribution $g(b) = q\ (1 - q)^b$, where B is the mean and $q = \frac{1}{B}$. Therefore, to model for single wavelength channel, $S^{(n)}$ is chosen to represent the state of the of the wavelength in any slot n. Noting that if the wavelength channel is occupied then $S^{(n)} = 1$, and if not occupied $S^{(n)} = 0$. Arrivals to one wavelength channel are represented by α, which is given by $\alpha = {}^{\lambda}/_{M}$, where *M* represents all wavelength channels on a link, and each wavelength channel can be represented as a non-homogeneous Markov chain with transitional probability matrix $Q^{(n)}$.

$$Q^{(n)} = \begin{bmatrix} P_{0,0}^{n} & P_{0,1}^{n} \\ P_{1,0}^{n} & P_{1,1}^{n} \end{bmatrix} \tag{6.4}$$

Whereby the transitional probabilities inside the matrix $Q^{(n)}$ are given as follows:

$$P_{i,j}^{n} = P\ (S^{n+1} = j | S^{n} = i) \qquad\qquad i,j = 0,1 \tag{6.5}$$

To determine the number of reserved channels on a link output, $Z^{(n)}$ is used to represent reserved channels in any slot *n*, therefore $Z^{(n)}$ = 0, 1, 2, …*M*. Likewise, the output link can also be modelled as a non-homogeneous Markov chain with transitional probability matrix $X^{(n)}$.

$$X^{(n)} = \begin{bmatrix} x_{0,0}^{n} & x_{0.1}^{n} & \dots x_{0,M}^{n} \\ x_{1,0}^{n} & x_{1.1}^{n} & \dots . x_{1,M}^{n} \\ x_{M,0}^{n} & M_{M,1}^{n} & \dots x_{M,M}^{n} \end{bmatrix}$$

Whereby the transitional probabilities inside the matrix $X^{(n)}$ are given as follows:

$$X_{i,j}^{(n)} = P\left(Z^{n+1} + j \middle| Z^{(n)} = i\right) \qquad i = 0, 1, 2, \dots M \tag{6.6}$$

Otherwise, the stationary probability for $Z^{(n)}$ is given as follows:

$$\Pi = (\Pi_0, \Pi_{1,} \ \dots \Pi_M)$$

Where Π can be solve using equation 6.7:

$$\Pi_j = \sum_{i=0}^{M} x_{i,j}^{o} . \Pi_i \qquad 0 \leq j \leq N \tag{6.7}$$

Using the mentioned parameters, then the probability of blocking can be calculated to a more suitable approximation because all variables for an OBS-JET network are considered in the analysis. Therefore, to calculate a blocking probability of a burst, a reference data burst with duration b slots is considered, offset time a slots, and wavelength state for all slots $n = a, \ a + 1, \dots a + b - 1$. Furthermore let the probability of $Z^{(n)}$ be:

$$V_z^{(n)} = P\left(Z^{(n)} = z\right) \qquad z = 0, 1, 2, \dots M \tag{6.8}$$

Then let $T_s(a, b)$ represent the probability initially at state s in time slot a, then $S^{(n)}$ will remain at state s for the duration of time slot b. Therefore, the blocking probability of a DB with offset time a and duration size b can be expressed as equation 6.9.

$$BP(a, b) = 1 - \ T_0(a, b) \sum_{z=0}^{M-1} V_z^{(a)} \tag{6.9}$$

Where $V_z^{(a)}$ can be solved as:

$$V_z^{(a)} = \Pi X^0 X^{(1)} \dots X^{(a-2)} X^{(a-1)} \tag{6.10}$$

Where $X^{(a-1)}$ represents the Z^{th} matrix $X^{(a-1)}$. Basically the equation in 6.9 for the blocking probability of a DB implies that the probability that a reservation request is accepted is equal to the probability that one or more wavelength channels are unused at the time the DB arrives and will further remain unused for the time period of the DB length or longer. The probability that the wavelength will remain free for the whole duration of DB length can be represented as follows:

$$T_0(a, b) = P_{0,0}^{(a)} P_{0,0}^{(a+1)} \dots . P_{0,0}^{(a+b-2)}$$

$$T_0(a,b) = e^{-\alpha(a+1)} e^{-\alpha(a+2)} \dots e^{-\alpha(a+b-1)} \tag{6.11}$$

Noting that if the burst offset time constitute a uniform distribution, therefore equation 6.3 can be re-written as:

$$\lambda_{(n)} = \begin{cases} \lambda, & for\ n \le 0 \\ \lambda \left[1 - \frac{n}{1 + A_{max}}\right], & for\ 0 < n \le A_{max} \\ 0, & for\ n > A_{max} \end{cases} \tag{6.12}$$

Where, A_{max} is the maximum offset duration.

6.6.2 Model Proof of Concept and Performance Evaluation

Performance evaluation of the model for the probability of blocking in OBS-JET networks is analyzed for different burst offset size a and different burst duration size b. Assuming a maximum burst offset time (A_{max}) of 180 µs, burst duration b = 81920 bits and slot size β = 20480 bits, and the data channel bandwidth is 2 377 728 000 bps. Therefore, the slot size β (in seconds) = 8.6 µs, and the burst duration b (in seconds) = 34.4 µs. The blocking probability is evaluated based on the mentioned parameters and Figure 6.11 shows the corresponding results with different offset time a for the analytical model approximation.

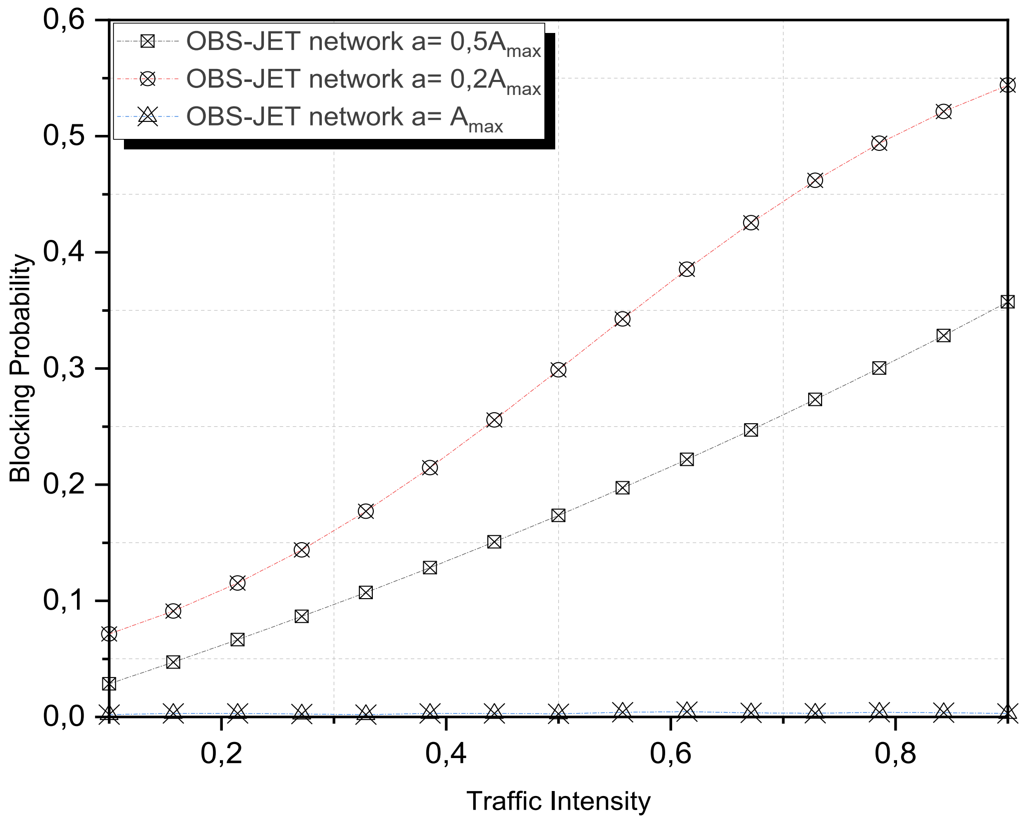


Figure 6. 11 Effects of the offset size (a) on OBS-JET networks

What can be observed in figure 6.11 is that if the offset time increases, the blocking probability reduces which implies that $BP\ (0.2\ A_{max}) > BP\ (0.5\ A_{max}) > BP\ (A_{max})$. This is an acceptable trend because the larger the offset time implies that the requested interval becomes even further away, hence less number of requests will overlap with each other, but noting that if the offset time becomes too large that can negatively impact on the latency requirements. Finally, it can also be observed that the probability of blocking increases proportional to the traffic intensity.

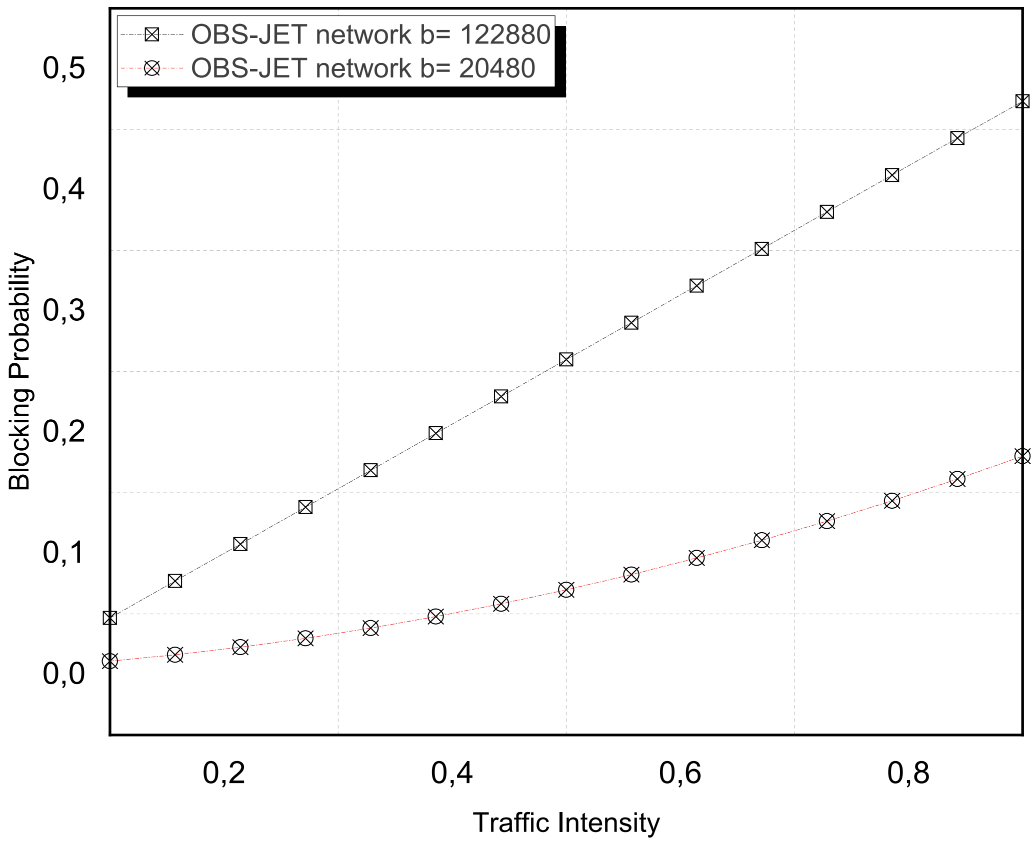


Figure 6. 12 Effects of the burst length on OBS-JET networks

Figure 6.12 shows the results with different burst length or size, b_1 = 20480 bits which is equivalent to 34.4 µs, and b_2 =122880 bits which is equivalent to 51.7 µs, while the offset time is constant at $0.5A_{max}$ = 90µs and the slot size β =5120 bits which is equivalent to 2.2µs. It can be observed in Figure 6.12 that the blocking probability decreases if the burst size decreases. This is no surprise because the burst duration b affects the probability that the wavelength scheduler will successfully locate a wide enough gap on the wavelength channel to fit the incoming burst, hence the bigger the burst size, the less chance of locating a suitable gap, resulting in a higher blocking probability because of a larger value of b.

6.7 Evaluation of Elastic Optical Transport Network

As mentioned in chapter 4, elastic optical transport networks play a significant role in achieving and maintaining energy efficiency of the entire network as compared to other traditional fixed-grid approaches. The architecture in Figure 4.8 represents the building blocks of an elastic network, and Figure 6.6 represents an elastic design of the entire network in modern networks. Furthermore, algorithm 2 represents physical layer impairment aware routing. Therefore, to evaluate the efficacy of the PLI-aware RWA in elastic networks a blocking probability is compared for traditional RWA in WDM networks against PLI-aware RWA in elastic networks. This is represented in Figure 6.13. It should be noted that in this scenario, a network topology with 14 nodes is used. This network comprises of 22 bidirectional fiber links between nodes. Following algorithm 2 and assuming the entire network uses a SMF consisting of an attenuation constant α =0.25 dB per km, fiber dispersion D =11.9 ps/nm/km, and non-linearity constant γ = 1.49 W/km. Finally, 38 wavelength channels per link are assumed.

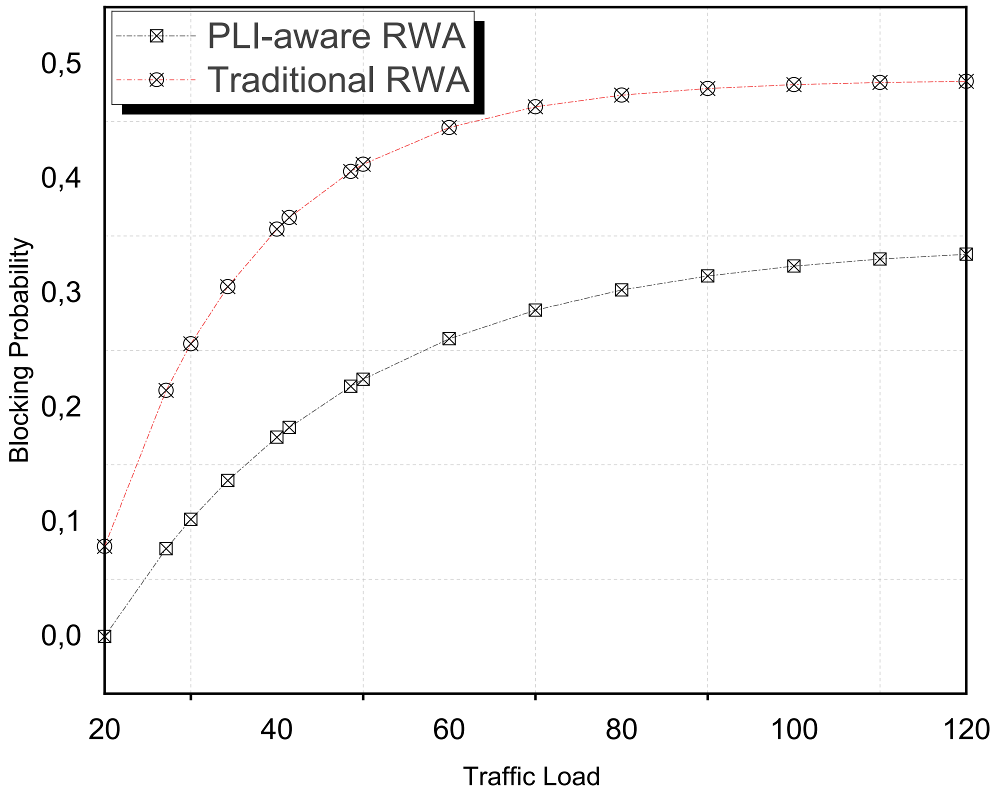


Figure 6. 13 The blocking probability of PLI-aware RWA vs traditional RWA

From the comparison in Figure 6.13, it can be observed that the blocking probability of PLI-aware RWA is much lower than traditional RWA, this is no surprise because the PLI-aware RWA considers physical layer impairments on each light path. Furthermore, An EON based on OFDM is compared with various networks operation approaches such as SLR (10G/40G/100G) and MLR as illustrated in Figure 6.14.

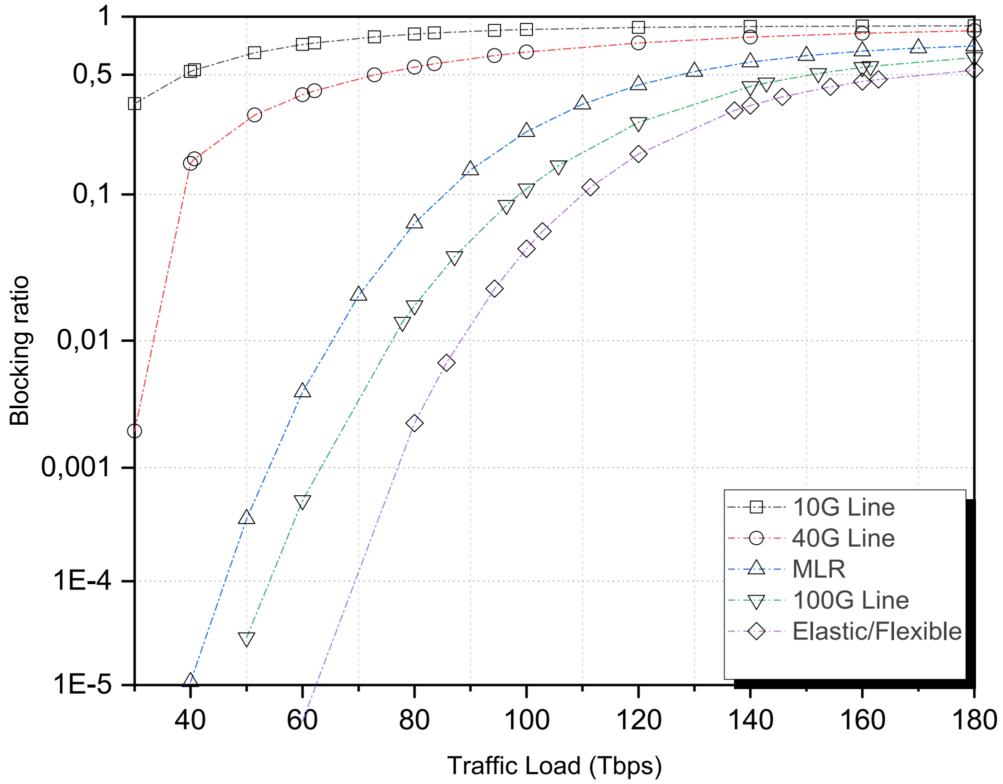


Figure 6. 14 The blocking probability comparison between various network approaches

The comparison in Figure 6.14 shows that the elastic or flexible optical transport network has comparably the lowest service blocking than other network operations (SLR 10G/40G/100G and MLR). Elastic optical transport network is clearly outperforming all other network operations, this is due to its ability to adjust and adapt to the actual demanded transmission rate hence being more energy efficient than other network operations.

6.8 Summary Chapter Conclusions

In this chapter, we discussed the desirable design and operational features for transport networks, which include architectural design and operation towards energy efficiency. We also mitigate approaches towards achieving an elasticity in the architecture based on which a framework for an advanced network approach that can efficiently process a large amount of data in the fastest time possible is presented. It is termed herein as a joint all photonic and wireless transport network architecture. The architecture is demonstrated in its capabilities of servicing high-capacity mobile backhaul and front-haul traffic, as well as real-time services support. Both candidate routing switching methods in the network are also explored. At the routing level, the work proposes a PLI-aware based RWA algorithm in the optical sections of the joint network, coupled with load-aware sleep-mode operation in the wireless sections. OBS is explored as a candidate solution switching paradigm in the core network. We opted for an elasticity in allocation in the overall optical transport network at the resource allocation level.

7. Conclusion and Future Work

In this work, we investigated energy-efficient approaches for both optical and wireless transport networks. The focus is mainly on architectural design and resource allocation. We explored elastic OTNs in view of achieving energy efficiency in their operation, It was ascertained that in this regard, they outperform their non-elastic equivalents. We also discussed the incorporation of SDN/NFV in future generation network deployments. SDN/NFV will indeed improve network management and control and at the same time minimize OPEX and CAPEX expenditures. Physical impairments hamper overall optical network performance in that they decrease overall optical transmission reach, and this prompted us to also mitigate appropriate regenerator placement approaches that would balance between overall transmission reach versus energy efficiency. We ultimately proposed an energy-efficient regenerator placement strategy based on OSNR, which thrives in keeping the number of regenerator nodes and OPEX thereof to a minimum whilst at the same time maintaining acceptable QoT levels. We further discussed the desirable design and operational features for transport networks, which include architectural design and operation towards energy efficiency. We also mitigate approaches towards achieving elasticity in the architecture on the basis of which a framework for an advanced network approach that can efficiently process a large amount of data in the fastest time possible is presented. It is termed herein as a joint all photonic and wireless transport network architecture. The architecture is demonstrated in its capabilities of servicing high-capacity mobile backhaul and front-haul traffic, as well as real-time services support. Both candidate routing and switching methods in the network are also explored. At the routing level, the work proposes a PLI-aware based RWA algorithm in the optical sections of the joint network, coupled with load-aware sleep-mode operation in the wireless sections. OBS is explored as a candidate solution switching paradigm in the core network. At the resource allocation level, we opted for elasticity in allocation in the overall optical transport network.

In the future, we will explore ways of enhancing scalability in all photonic networks by introducing optical cell switching (OCS). This is because currently all active wavelengths are assumed to be switched independently as we route data across the network. We will further investigate synchronization issues in OCS. Recall that OCS essentially divides time into fixed-size slots and then bundles several wavelengths in a single slot. Synchronization issues arise at the switching nodes. Thus we will explore both the aligned versus non-aligned operations of OCS in regards to synchronization at the switching nodes.

References

[1] Y. Zhang, P. Chowdhury, M. Tornatore, and B. Mukherjee, "Energy Efficiency in Telecom Optical Networks,"IEEE Communications Surveys & Tutorials, vol.12, no.4, pp. 441-458, Nov. 2010.

[2] C. Gabriel, A. He, and A. Chern, "Green 5G: Building a Sustainable World, "Analysys Mason, July 2020.

[3] M. Fiorani, B. Skubic, J. Martensson, L. Valcarenghi, P. Castoldi, L. Wosinska, and P. Monti, "On the design of 5G transport networks,"Photon Network Communication (2015), vol. 30, pp. 403-415.

[4] Ericsson, "Ericsson Mobility Report," http://www.ericsson.com/res/docs/2014/ericsson-mobility-report-june- 2014.pdf, June 2014.

[5] S. Han, and S. Bian, "Energy efficient 5G for a greener future," Nature Electron, vol. 3, no. 4, pp. 182-184, April 2020.

[6] M.A. Al-Namari, A. Mansoor, and M.Y. Idris, "A brief Survey on 5G wireless mobile networks," International Journal Adv. Comput. Sci. Appl., Vol. 8, no. 11, pp. 52-59, 2017.

[7] A. Mughees, M. Tahir, M.A. Sheikh, and A. Ahad, "Towards Energy Efficient 5G Networks Using Machine Learning: Taxonomy, Research Challenges, and Future Research Directions," IEEE Communication Access, Oct. 2020.

[8] A. Fonseca, R. Kazman, and P.Lago, "A manifesto for energy-aware software," IEEE softw., vol. 36, no. 6, pp. 79-82, Nov. 2019.

[9] X. Duan, A. M. Akhtar, X. Wang, "Software-defined networking-based resource management: Data offloading with load-balancing in 5G HetNet," Journal on Wireless Communication and Networking, vol. 12, no. 1, pp. 1-13, 2015.

[10] J. G. Andrews, S. Buzzi, "What will 5G be?," IEEE Areas Communication, 2014.

[11] H. Ishii, Y. Kishiyama, and H. Takahashi, "A novel Architecture for LTE-B: C-plane / u-plane split and Phantom Cell concept," pp. 5-25, IEEE Globecom Workshop, Dec 2012.

[12] D. Sinh, L. Tung, "The challenges for Applying SDN/NFV for 5G & IoT," Researchgate Conference Paper, pp. 3-9, August 2017.

[13] R. Sabella, A. Bianchi and G. Bottari, "Trends in optical transport networks and related technologies," IEEE Ericson research.

[14] S. Iyer, and S.P. Singh, "Physical Layer Impairment-Aware Routing and Wavelength Assignment (PLI-RWA) Strategy for Mixed Line Rate (MLR) Wavelength Division Multiplexing (WDM) Optical Networks," IEEE, 2015.

[15] J.L. Vizcaino, "Energy-Efficient Design of Optical Transport Networks," University Thesis, Dec 2016.

[16] M Arumugam, "Optical fiber communication—An overview," Fundamental Aspects of Solitons, Fibre Optics And Nonlinear Materials, November 2001.

[17] J. Cwlinski, "Introduction to Fiber Optics," Stony Point, New York.

[18] M. Kumari, R. Sharma "Passive Optical Network Evolution to Next Generation Passive Optical Network: A Review, " International Conference on Wireless Networks & Embedded Systems, 2018.

[19] Y. S. Kavian, "Resilient Optical Networks: Principles and State-of-the-Art," International Conference Communication Engineering, 24 Dec 2010.

[20] Y. Belayneh, "Sleep Mode Operation of Optical Networks for Power Consumption Minimization: a Case Study of Ethio Telecom Backbone Optical Transport Network," Master of Science Thesis, November 2018.

[21] O.D. Apuke, "Introduction to Radio Production," Lambert academic publishers, May 2017.

[22] H. Gupta, P. Gupta, P. Kumar, "Passive Optical Networks: Review and Road Ahead," Proceedings of TENCON, October 2018.

[23] A. Checko, "Cloud Radio Access Network architecture," PhD Thesis, Radiocomp MIT company, Denmark 2016.

[24] S. Zeadally, S.U. Khan, N. Chilamkurti, "Energy-Efficient Networking: Past, Present, and Future," Springer Science, 31 May 2011.

[25] M. Usama, and M. Erol-kantarci, "A survey on recent trends and open issues in energy efficiency of 5G," Sensors, vol. 11, no. 14, pp. 697-709, April 2016.

[26] A. A. Barakabitze, A. Ahmad, "5G network slicing using SDN and NFV: A Survey of taxonomy, architectures and future challenges," Computer networks, November 2019.

[27] A. Mughees, M. Tahir, and A. Ahad, "Towards energy efficient 5G networking using machine learning: A taxonomy, research challenges, and future research directions," IEEE Access, Sept 2020.

[28] O. Idowu-Bismark, O. Kennedy, R. Husbands, M. Adedokun, "5G Wireless Communication Network Architecture and Its Key Enabling Technologies," International Review of Aerospace Engineering (I.RE.AS.E), Vol. 12, N. 2, April 2019.

[29] P. Goransson, C. Black, "Software-Defined Networks: A comprehensive approach," Morgan Kaufmann, 2014.

[30] Y. Wang, S. You, "An efficient route management framework for load balance and overhead reduction in SDN-based data center networks," IEEE Transactions on networks and services managemen, vol. 15, no. 4, Dec. 2018.

[31] D.B. Hoang, M. Pham, "On software-defined networking and the design of SDN controllers," IEEE access, International Conference on the Network of the Future, Sept 2015.

[32] K.D. Ramos, F.M. Esteres Verissimo, "Software defined networking: A comprehensive survey," Proceedings of IEEE, 2015, vol. 103, pp. 14-74.

[33] Y. Wang, S. You, "An efficient route management framework for load balance and overhead reduction in SDN-based data center networks," IEEE Transactions on networks and services managemen, vol. 15, no. 4, Dec. 2018.

[34] A. Basta, A. Blenk, K. Hoffman, H. J. Morper, "Towards a cost optimum design for a 5G mobile core network based on SDN and NFV," IEEE Transactions on Network and Service Management, Vol. 14, No. 4, December 2017.

[35] A Basta, W. Kellerer, M. Hoffman and E. D. Schmidt, "A virtual SDN-enabled LTE EPC architecture: A case study for gateway functions," Proc. IEEE SDN future Networks, pp. 1-7, 2013.

[36] P. Mahadevan, P. Sharma, S. Banerjee, P. Ranganathan, "Energy Aware Network Operations," IEEE INFOCOM Workshops, 2009.

[37] N. W. Williams, "A traffic aware sleep mode strategy for energy efficient dense HETNETS," School of Engineering (SEng.), February 2021.

[38] F. Idzikowski, L. Chiavaviglio, A. Clanfrani, "A Survey on Energy-Aware Design and Operation of Core Networks," IEEE Communication Survey, Vol. 18, 2016.

[39] R. Bajpai, S. Iyer, S. P. Singh and S. Sengar, "Energy-efficient and spectral-efficient mixed line rate optical WDM networks: a comparison," Inder Science, April 2020.

[40] R. Agorwal, R. Bhatia, "A Review of Mixed Line Rate and Elastic Optical Network," IEEE International Conference on Computing for Sustainable Global Development, India, 2019.

[41] P. Chowdhury, M. Tornatore, A. Nag, T. Wang, "On the Design of Energy-Efficient Mixed-Line-Rate (MLR) Optical Networks," IEEE Journal of Lightwave Technology, November 2011

[42] S, Suthaharan, P. Rukshani, C. Vanuja, "Frequency spectrum analysis of mixed-line rate in flexible optical networks," Digital Research Repository, University of Jaffna, 2019.

[43] V. Rozite, "Network Standby Towards a Policy Framework," International Energy Agency, March 2013.

[44] T.R. Wollesen, "Energy Efficient Improvement in 5G Mobile Networks," Masters Thesis, Technical University of Denmark, June 2018.

[45] C. Han, "Green Radio: Radio Techniques to Enable Energy Efficient Wireless Network," IEEE Communication Magazine, vol. 49, 2011.

[46] G. Auer, O. Blume, "Earth Project D2.3-Energy Efficient Analysis of the Reference Systems, Areas of Improvements and Target Breakdown," Energy-aware Radio Network Technology Conference. 2012.

[47] T.R. Wollesen, "Energy efficiency improvements in 5G mobile network," Master's Thesis, Technical University of Denmark 2018.

[48] X. Chu, D.L. Perez, X. Yang, and F. Gunnarsson, "Heterogeneous Cellular Networks," Cambridge University Press, 2013.

[49] S. Rizvi, A. Aziz, M. T. Jilani, N. Armi, G. Muhammad, "An investigation of energy efficiency in 5G wireless networks," International Conference on Circuits, System and Simulation (ICCSS), 2017.

[50] J. Onyeka, A. Imoize and A. Atayero, "Energy efficient design techniques in next generation wireless communication networks: Emerging trends and future directions," Wireless communications and mobile computing, 2020.

[51] A. Vasjanov, V. Barzdenas, "A Review of Advanced CMOS RF Power Amplifier Architecture Trends for Low Power 5G Wireless Networks," Microwave and Wireless Communications, October 2018.

[52] I. Neokosmidis, T. Rokkas, P. Paglierani and P.S. Khodeshanas, "Assessment of Capex and Opex for Media Services in cloud enabled 5G networks," Smart Cities information communication Technology, pp. 1-6, Sept 2019.

[53] A. Checko, H. L. Christiansen, Y. Yan, L. Scolari, G. Kardaras, "Cloud RAN for Mobile Networks - a Technology Overview," IEEE Communications Surveys and Tutorials, 2014.

[54] N. Hassan, K.A. Yau, C. Wu, "Edge Computing in 5G: A Review," IEEE Access, August 2019.

[55] N. Abbas, Y. Zhang, A.Taheckordi, and T. Skeie, "Mobile Edge Computing: A survey," IEEE internet things, vol. 5, no.1, pp. 450-465, Feb 2018.

[56] K. Zhang, Y. Mao, S. Leng and Q. Zhao, "Energy Efficient Offloading for Mobile Edge Computing in 5G Heterogeneous Networks," Green Communications and Networking for 5G Wireless, IEEE Access, Aug 2016.

[57] M. Fiorani, B. Skubic, "On the Design of 5G Transport Network," Photon Network Communication, pp. 403-415, 2015.

[58] S. Chouhan, D. K. Panda, M. Gupta, S. Singhal, "Multiport MIMO antennas with mutual coupling reduction techniques for modern wireless transreceive operations: A review," International Journal of RF and Microwave Computer-Aided Engineering, 2018.

[59] L. Lu, G.Y. Li, A.L. Swandlehurst, "An Overview of Massive MIMO: Benefits and Challenges," IEEE Journal in Signal Processing, Vol. 8, October 2014.

[60] T. L. Marzatta, "Massive MIMO: An Introduction," Bell Labs International Journal, Vol. 20, 2015.

[61] D. Feng, L. Lu, Y. Y. Wu, G. Y. Li, G. Feng, " Device-to-Device communication in Cellular Networks," IEEE Communication Magazine, April 2014.

[62] W. J. Jung, H. J. Ki, T. J. Lee, M.Y. Chung, "Adaptive Sleep Mode Algorithm in IEEE 802.16e," Proceedings of Asia-Pacific Conference on Communication, 2007.

[63] F. Idzikowski, E. Bonetto, L. Chiavaviglio, A. Coiro, R. Duque, J. Lopez Vizcalno, "TREND in Energy-Aware Adaptive Routing Solutions," IEEE Communication Magazine, November 2013.

[64] X. Weng, D. Cao, Z. Niu, "Energy- Efficient cellular Network Planning Under Insufficient Cell Zooming," IEEE Technology Conference, Vol. 38, 2011.

[65] T. Alviola, "Energy efficiency in wireless networks." Master's Thesis in Information Technology, September 2013.

[66] F. Salahdine, J, Opadere, Q Liu, T. Han, N. Zhang, "A Survey on Sleep Mode Techniques for Ultra Dense Networks in 5G and Beyond," Computer Networks, December 2021

[67] M. Masoud, M.G. Khafegy, E. Soroush, and D. Giacomelli, "Reinforcement Learning for Traffic-Adaptive Sleep Mode Management in 5G Networks," IEEE Annual International Conference, August 2020.

[68] A. Sinha, A. Chandrakasan, "Dynamic Power Management in Wireless Sensor Networks," IEEE Design and Test of Computers, March 2001.

[69] Dr. S. Sasipriya, R. Vigneshram, "An Overview of Cognitive Radio in 5G Wireless Communication," International Conference on Computational Intelligence and Computing Research, IEEE Access, Dec. 2016.

[70] S. Iyer, and S.P. Singh, "Spectral and power efficiency investigation in single- and multi-line-rate optical wavelength division multiplexed (WDM) networks," Photon network communication, 2017.

[71] A. Frontinele, I. Santos, D.R. Campelo, A. Soares, "An efficient IA-RMLSA Algorithm for transparent elastic optical networks," Computer Networks, pp. 1-14, March 2017.

[72] G. Zhang, M. D. Leenheer, A. Morea, B. Mukherjee, "A Survey on OFDM-Based Elastic Core Optical Networking," IEEE Communications Surveys & Tutorials, February 2012.

[73] B. Mukherjee, "Optical WDM Networks," Springer Science, pp. 1-39, 2006

[74] I. Tomkos, E. Palkopoulou, M. Angelou, “A survey of recent developments on flexible/elastic optical networking,” International Conference on Transparent Optical Networks (ICTON), July 2012.

[75] Y. Tan, R. Gu, Y. Ji, “Energy-efficient routing, modulation and spectrum allocation in elastic optical networks,” Optical fiber technology, 2017.

[76] H. Liu, Q. Xiong, And Y. Chen, “Routing Core and Spectrum Allocation Algorithm for Inter-Core Crosstalk and Energy Efficiency in Space Division Multiplexing Elastic Optical Networks,”IEEE , April 2020.

[77] R. Krishnamurthy, T. Srinivas, “Physical Layer Impairment Aware Routing and Spectrum Allocation Algorith for Transparent Flexible-grid Optical Networks,” Computer Communications, pp. 507-514, 2020.

[78] J.L. Vizcaino, Y. Ye, I. T. Monroy, “Energy efficient analysis for flexible-grid OFDM-based optical networks,” Computer Networks, 2012.

[79] X. Chen, B. Guo, J. Guan, S. Ma, J. Li, Z. Chen, Y. He, “Optimized hierarchical OXC architecture for elastic optical network,” IEEE, July 2014.

[80] K. Akaki, P. Pavarangkoon, N. Kitsuwan, “Elastic optical network for fragmented bandwidth allocation with limited slicers,” International Conference on Information Technology and Electrical Engineering, 2021.

[81] H. Takara, M. Jinno, “Experimental Demonstration of Spectrum-Sliced Elastic Optical Network (SLICE),” Optics express, Oct 2010.

[82] S. Dong, “Regenerator Placement and Fault Management in Multi-Wavelength Optical Networks,” MSc thesis, July 2011.

[83] A. Kahya, “Routing, Spectrum Allocation and Regenerator Placement in Flexible-Grid Optical Networks,” MSc thesis, August 2013.

[84] I. Nath, M. Chatterjee, U. Bhattacharya, “A survey on regenerator Placement Problem in translucent optical network,” IEEE International Conference on Circuits,” April 2014.

[85] X. Yang, B. Ramamurthy, “Sparse Regenerator in Translucent Wavelength-Routed Optical Networks: Architecture, Network Design and Wavelength Routing,” Photonic Network Communication, pp. 39-53, March 2005.

[86] M. Attaran, “The impact of 5G on the evolution of intelligent automation and industry digitization,” Journal of Ambient intelligence and Humanized Computing, Sept 2020.

[87] A. Srivastava, M.S. Gupta, and G. Kaur, “Energy efficient transmission trends towards future green cognitive radio networks (5G): Progress, taxonomy and open challenges,” Journal of Network and Computer Applications, July 2020.

[88] Ericsson Mobility Report Mobile Data Traffic Outlook, https://www.ericsson.com/en/mobility-report/dataforecasts/mobile-traffic-forecast. (accessed Dec. 3, 2021).

[89] A. Mirzaeinnia, M. Mirzaeinia, and A. Rezgui, “Latency and throughput optimization in modern networks: A comprehensive survey,” IEEE comm. Survey and Tutorials Journal.

[90] S. Banerjee, A. Ghosh, S. Mahapatra, and R. Karmakar, “Performance survey of MANET routing protocols with TCP congestion control algorithms,” IEEE International Conference for Convergence in Engineering, 2020.

[91] D.K.N. Venkatramara, S.B. Srikantaiah, and J. Moodabidri, “SCGRP: SDN-enabled connectivity-aware geographical-aware routing protocol of VANETS for Urban environment,” IET Networks, pp. 102-111, 2017.

[92] W. Duan, Y. Ji, Y. Zhang, G. Zhang, V. Frascolla and X. Li, “5G Technology Based Remote E-Health: Architecture, Applications, and Solutions,” Sept 2020.

[93] M. A. Chalouf, “Managing Quality of service and security in an e-Health environment,” Service Level Management in Emerging Environments, 2020.

[94] Z. Guizani, N. Hamdi, “CRAN, H-CRAN, and F-RAN for 5G systems: Key capabilities and recent advances,” Wiley, March 2017.

[95] R. Krishnamurthy, T. Srinivas, “Physical layer impairments aware routing and spectrum allocation algorithm for transparent flexible-grid optical networks,” Computer Communications, 2020, pp. 507–514.

[96] T. Takagi, H. Hasegawa, K. Sato, T. Tanaka, B. Kozicki, Y. Sone, M. Jinno, Algorithms for maximizing spectrum efficiency in elastic optical path networks that adopt distance adaptive modulation, in: Proc. European Conf. on Optical Communication, Turin, Italy, 2010.

[97] K. Christodoulopoulos, I. Tomkos, E.A. Varvarigos, Routing and spectrum allocation in OFDM based optical networks with elastic bandwidth allocation, in: Global Telecommunications Conference (GLOBECOM 2010), 2010 IEEE, IEEE, 2010, pp. 1–6.

[98] A.G. Rahul Kumar, Flexible Grid Optical Network Architectures: Key Issues and Possible Solutions, M.Tech, project report, Indian Institute of Technology, Bomay, 2013.