Flexible Optical Systems and Reconfigurable Disaggregated Data Centre Networks

Adaranijo Peters

Department of Electrical and Electronic Engineering

University of Bristol

A dissertation submitted to the University of Bristol in accordance with the requirements for award of the degree of *Doctor of Philosophy* in the Faculty of Engineering.

May 2018

“In memory of my beloved mother”

Abstract

Optical metro/core and data centre networks are under pressure to manage the exponential growth and dynamic nature of internet traffic. As a result, innovative technologies and approaches to improve the flexibility of the network infrastructure and operations need to be developed. However, the provisioning of networks with innovative technologies and resources to deliver flexibility without any design or allocation policy does not guarantee cost efficiency and optimum network performance when managing traffic demands. The introduction of flexibility as a measurable key performance indicator for networks has opened up potential prospects for the design of efficient optical networks. However, there is limited understanding on how quantitative levels of flexibility relates to other key performance indicators and design features. In addition, the combination of server resource disaggregation, electrical/optical technologies and network function programmability in data centres provides a promising solution to eliminate the limitations of conventional server-centric data centres. However, for this concept to materialize, resource allocation policies which ensure an optimum level of network performance must be developed.

The first part of this thesis investigates flexibility as a measurable key performance indicator for optical transmission and switching systems. Flexibility measurement models for different optical components and subsystems are derived and proposed based on maximum entropy. In addition, the flexibility of the examined optical components and subsystem are measured, and the relationship between flexibility, other measurable key performance indicators and design features are highlighted and discussed. The second part of this thesis investigates network strategies and algorithms for selecting and deploying electronic packet switching/optical circuit switching services and network resources to build Virtual Machines on reconfigurable disaggregated data centre resources. A comprehensive performance evaluation of the proposed network strategies and algorithms across different disaggregated data centre architectures is conducted.

Acknowledgement

First and foremost, I will like to thank and appreciate my supervisor Dr Georgios Zervas for his support, guidance and advice throughout the duration of my PhD programme. Dr Zervas is a hardworking and technically sound academic, he always motivated and challenged me to produce high quality research output. I am also grateful to him for giving me the opportunity to join his research team on the dRedBox project.

I will also like to thank my co-supervisor Dr George Oikinomou for accepting to be my supervisor during the later stages of my PhD program. His sound advice and dedication put me on track to the completion of my PhD program.

I will like to thank Dr Emilio Hugues-Salas for his input in my PhD research. Emilio was always available to assist me anytime I had problems in any aspect of optical networks. I will also like to appreciate my fellow PhD colleagues Hui, Rui and Kostas for our productive research discussions.

Special thanks to my girlfriend Omotola for her support, encouragement, love and care throughout the duration of my PhD program. I also appreciate my friends Deola, Aba, Eya and Tesleem for their support.

Special thanks and appreciation to my Dad, Professor Olufemi Peters for his encouragement and financial support which assisted me through my PhD program. I also appreciate Aunt Wemi and my siblings Dolapo, Damilola and Kehinde for their encouragement and support.

I am also grateful to National Open University of Nigeria for providing me with financial support.

Declaration

I declare that the work in this dissertation was carried out in accordance with the requirements of the University's Regulations and Code of Practice for Research Degree Programmes and that it has not been submitted for any other academic award. Except where indicated by specific reference in the text, the work is the candidate's own work. Work done in collaboration with, or with the assistance of, others, is indicated as such. Any views expressed in the dissertation are those of the author.

Adaranijo Peters

May 2018

Table of contents

List of Figures

List of Acronyms

AoD	Architecture on Demand
AWGR	Arrayed Waveguide Grating Routers
BS	Broadcast and Select
BVT	Bandwidth Variable Transponder
CDC	Colourless, Directionless and Contentionless
DC	Data Centre
D-COS	dRedBox Congestion Aware Placement Strategy
D-NES	dRedBox Network-Hop Aware Placement Strategy
D-RAS	dRedBox Random Placement Strategy
dRedBox	Disaggregated Recursive Data Centre in a Box
EON	Elastic Optical Network
EoT	Edge of Tray
EPS	Electronic Packet Switching
HBTP	High Bandwidth Traffic Pattern
I/O	Input/Output
KPI	Key Performance Indicator
LBTP	Low Bandwidth Traffic Pattern
MCS	Multicast Switch
MPSoC	Multi-Processor System on Chip
NFV	Network Function Virtualisation
RDC-1	Heterogeneous Racks with Stacks of Heterogeneous Trays Category
RDC-2	Heterogeneous Racks with Stacks of Homogeneous Trays Category
RDC-3	Homogeneous Racks with Stacks of Homogeneous Trays Category
ROADM	Reconfigurable Add/Drop Multiplexer
OCS	Optical Circuit Switching

SBVT	Sliceable Bandwidth Variable Transponder
SDN	Software Defined Networking
SR	Spectrum Routing
SSS	Spectrum Selective Switch
SSDF	Switch and select with dynamic functionality
ToC	Top of Cluster
ToR	Top of Rack
TWC	Tunable Wavelength Converters
WDM	Wavelength Division Multiplexing
WSS	Wavelength Selective Switch
VNF	Virtual Network Function
VM	Virtual Machine
1-Tier-H	1-Tier Hybrid Disaggregated Data Centre Architecture
3-Tier-H	3-Tier Hybrid Disaggregated Data Centre Architecture
1-T-COS	1-Tier-H Congestion Aware Placement Strategy
1-T-NES	1-Tier-H Network-Hop Aware Placement Strategy
1-T-RAS	1-Tier-H Random Placement Strategy
3-T-COS	3-Tier-H Congestion Aware Placement Strategy
3-T-NES	3-Tier-H Network-Hop Aware Placement Strategy
3-T-RAS	3-Tier-H Random Placement Strategy

1 Introduction

1.1 Motivation and Problem Statement

The Cisco forecast in [1] reported that global internet traffic is estimated to rise in next few years and by 2021 will be about 127 times the total size of global internet traffic in 2005. The growth in internet traffic is attributed to an increase in demand from consumers for bandwidth-intensive internet applications such as internet video and online gaming. The global internet traffic growth will also lead to unpredictable traffic patterns due to dynamic consumer demands and behaviour. These traffic trends will have a direct impact on the resources of existing optical metro/core and data centre (DC) network infrastructure because these bandwidth-intensive applications are hosted on servers in DCs and are transmitted over optical metro/core networks between DCs and from DCs to consumers. According to [2], a major amount of internet traffic will be dominated by global DC traffic, and since 2008 the DC has been the origin or destination of most of internet traffic. It was also reported in [2] that by 2021, 14.9% of the global DC traffic will occur between DC to user, 13.6% of the global DC traffic will occur between DC to DC, and 71.5% of the global DC traffic with occur within the DC (Figure 1.1 illustrates an overview of flow of DC traffic across optical metro/core and DC network infrastructures). To efficiently manage these traffic trends, innovative and efficient technologies, network infrastructure and network resource planning policies for metro/core and DC networks must be developed to provide optimum levels of network flexibility and performance.

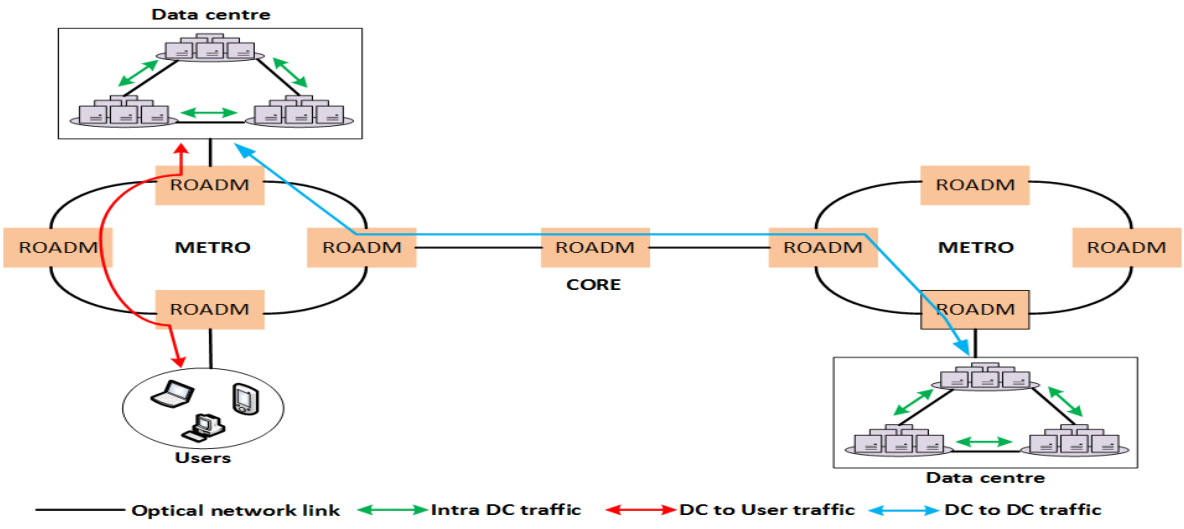

Figure 1.1: Optical metro/core and data centre network.

Due to the dynamic and diverse nature of internet traffic, flexibility is an essential prerequisite and key performance indicator (KPI) for the design of optical metro/core and DC networks. Researchers have explored various ways to deliver flexibility in networks with the development of innovative hardware technologies, network infrastructure, and software technologies such as software defined networking (SDN) and network function virtualisation (NFV) which enable network programmability and control. However, provisioning of networks with innovative technologies and resources to deliver flexibility without any design or allocation policy does not guarantee cost efficiency or optimum network performance in handling traffic demands. There two possible approaches to address this issue. The first approach is through developing efficient resource allocation strategies and algorithms that guarantee cost efficiency and optimum network performance. The second approach is through determining and quantifying different levels of flexibility required to deliver cost efficiency and ranges of certain performances. There have been extensive studies which have evaluated the flexibility of communication networks in respect to hardware technologies, network infrastructure and software technologies, however, the performance of these approaches were evaluated in terms of KPIs such as blocking probability, cost and energy efficiency, and not as a function of flexibility as a measurable KPI. Interestingly, in the context of manufacturing, an extensive amount of studies have examined flexibility as a measurable KPI and proposed several measurement techniques to quantify flexibility [3]. However, in the context of communication networks, flexibility as a measurable KPI has remained generally unexplored by research community, only studies in [4]–[7] have attempted to quantify flexibility as a measurable KPI. This can be attributed to the fact that flexibility can be defined in a wide variety of contexts or there is complexity involved in developing measurement models for the different building blocks of communication networks. The study of flexibility as a measurable KPI and the development of efficient resource allocation policies are key research areas in which continuous innovation is required to deliver flexible optical metro/core and DC networks with optimum network performance.

1.1.1 Optical Metro/Core Networks

In optical metro/core networks, one of the most significant elements introduced to add flexibility is the reconfigurable optical add/drop multiplexer (ROADM). ROADMs allow network operators to remotely select, switch and route wavelength channels, and also configure the network to support different traffic requirements without the need to physically disrupt

network operations [8]. Also, enabling technologies such as wavelength selective switch (WSS), multicast switch (MCS), couplers and optical transmitters that make up a ROADM's structure play an important role in defining the flexibility, functionality and performance of ROADMs. Therefore, the design features of each of these devices are vital for delivering efficient network services which can support unpredictable traffic trends. In addition, elastic optical networks (EONs) have emerged as a potential solution to address the limitations associated with fixed grid networks and meet the ITU G.694 grid channel spacing recommendation [9]. EONs provide flexible and efficient use of the spectrum to support different combinations of optical paths with variable bandwidths. The introduction of EONs have led to the development of devices such as spectrum selective switch (SSS) and bandwidth variables transponder (BVT) which can support flexible grid optical networks [10][11].

Regardless of prior efforts to deliver flexible optical networks, there are still critical challenges and questions that need to be addressed to effectively determine the levels of flexibility required to deliver cost effective and optimum network services. Therefore, crucial questions such as "what amount of flexibility in an optical component is required to achieve a certain level of performance?" are vital to understand how different levels of flexibility relates to design features, other KPIs and network resource requirements. This may assist network designers and equipment manufacturers to understand the amount of flexibility required to ascertain different levels of performance, which may in turn lead to resource savings, and a reduction in system complexity and equipment manufacturing time. In [4], a measurement approach to quantify flexibility of optical network elements based on maximum entropy was proposed. The authors proposed flexibility measurement models for different optical node architectures and switching optical components, and evaluated the flexibility of the different optical node architectures. The same flexibility measurement technique was also used in [5] to evaluate the flexibility of different add/drop bank structures considering the drop direction. Despite the previous works on flexibility measurement for optical networks and the promising potential of flexibility as a measurable KPI, there still remains an extensive range of complex optical switching and transmission components, and optical subsystems that have not been studied.

1.1.2 Data Centre Networks

DC networks have an important role to play in effectively managing the increasing and unpredictable traffic trend as bandwidth intensive applications such as internet video are hosted

and ran on DC servers and network infrastructure. With about 71.5% of DC traffic remaining within the DC [2], and conventional server-centric DCs infrastructure experiencing inefficient server utilisation, high-power consumption and cost-intensive scalability. Novel and cost-effective DC technologies, architectures, and resource allocation strategies must be investigated to deliver optimum levels of flexibility and network performance while managing the exponential growth and dynamic nature of internet traffic. Disaggregation of server resources into separate standalone CPU, memory, storage and network resource pools has been projected in [12][13] as a promising solution to improve resource utilization, efficiency, flexibility and upgradeability in DCs. In addition, hybrid DC architectures employing both electronic/optical technologies [14]–[16] and all-optical technologies [17]–[19] have been explored and have demonstrated benefits in terms of power consumption, modularity and cost in comparison to conventional DCs. Therefore, the combination of server resource disaggregation with hybrid technologies provides a promising solution to mitigate the drawback experienced by conventional server-centric DCs. However, for this concept to materialize, there are several challenges that need to be addressed. First, the authors in [12] stated that communication network will be a challenge for the realization of server resource disaggregation. This is because communication between resources (e.g. CPU to CPU and CPU to memory) for Virtual Machines (VM) placement which occurred within a server will now be distributed throughout the DC network. Secondly, existing hybrid DCs still suffer from one main drawback. The network function services in these hybrid DCs are hardwired with fixed electronic packet switching (EPS) to optical circuit switching (OCS) proportionalities and are dedicated to specific network resources. This limits the ability to dynamically deploy or upgrade network function services, which in turn limits the flexibility and performance of the DC to efficiently manage diverse network traffic requirements.

The dRedBox (Disaggregated Recursive Data Centre in a Box) architecture offers promising solutions to overcome these challenges. This architecture combines server resource disaggregation with state of the art software, optical and electronic technologies to deliver a DC architecture with efficient resource utilization, low power consumption, low latency, high throughput and non-disruptive upgradeability [13]. In addition, the dRedBox architecture provides improved flexibility and connectivity compared to conventional hybrid DC architectures due to deep multi-layer network function service programmability at end nodes (i.e., embedded CPU, memory) [20]. Unlike conventional hybrid DCs where a fixed amount of network resources for server input/output (I/O) ports are dedicated to either optical circuit

or electronic packet network, the dRedBox architecture allows for dynamic and run-time deployment of on-chip packet/circuit switching service to any I/O port to handle variable network requirements. Thus, VM requests can be implemented using a custom-built network topology with a network service chain spanning over multiple network hops. However, the introduction of these amount of flexibility and programmability features opens up challenges on how to effectively deploy and manage these network function services to deliver optimum network performance when deploying VMs. Typically, a VM request specifies the amount of CPU cores, memory and bandwidth connectivity required. Furthermore, the deployment of a VM request in disaggregated DCs requires two main phases: IT resource and network resource allocation. Factors such as resource availability, resource capacity, location of the different resources types across the DC network (i.e., how CPU and memory resources are arranged), network fabrics and allocation strategies should be considered when deploying VMs. Several studies [21]–[24] have proposed IT and network resource allocation algorithms for the deployment of VMs in disaggregated DCs and carried out simulation studies to evaluate the performance of the proposed algorithms. However, none of these studies implemented a comprehensive performance analysis or proposed policies and algorithms to address the challenges associated with the deployment and allocation of EPS/OCS network services to build VMs on reconfigurable disaggregated DC resources. The term "reconfigurable" and "programmable" are used interchangeably in this thesis.

1.2 Research Objectives and Contributions

The first research objective of thesis is to derive and propose flexibility measurement models for different design configurations of optical and transmission systems for flexible optical networks, and evaluate the design trade-offs and relationships between flexibility, measurable KPIs (i.e., capacity, connectivity, spectral efficiency, cost etc.) and design features. The research contributions associated with this research objective are:

- Derived and proposed models to measure the flexibility of WSS and SSS under different design conditions which include: WSS and SSS with no contention, WSS and SSS with contention, and WSS and SSS with port dimension reconfigurability.

- Derived and proposed models to measure the flexibility of different optical transmitter configurations. These transmitter configurations range from single carrier transmitters to multicarrier transmitters with fixed or variable transmission features such as

modulation formats, symbol rates, tuneable wavelength channels, electrical subcarriers and optical carriers.

- Derived and proposed models to measure the flexibility of optical subsystems with variable transmission and switching features. These subsystems include: WSS combined with BVTs, MCS combined with BVTs and SSS combined with BVTs.

- The flexibility of different configurations of the studied optical components and subsystems are measured using the proposed flexibility measurement models. Furthermore, the relationship between flexibility, other KPIs and design features are evaluated, and design trade-offs of the optical components and subsystems based on the results presented and theoretical analysis were highlighted.

The second research objective of this thesis is to develop and propose networking strategies and algorithms to select and deploy EPS/OCS services, generate multi-layer custom network topologies, and allocate network resources to builds VMs across the dRedBox architecture and other conventional hybrid disaggregated DC architectures in the most efficient way. The research contributions associated with this research objective are:

- Developed and proposed several networking strategies and algorithms for selecting EPS/OCS services, creating multi-layer custom network topologies and, allocating network resources to build VM network requests across the dRedBox architecture and other conventional hybrid disaggregated DC architectures.

- A comprehensive performance evaluation of the different networking strategies across the dRedBox and conventional hybrid disaggregated DC architectures is conducted and analysed under different traffic patterns.

- A benchmark of the dRedBox architecture against conventional hybrid DC architectures in terms of blocking probability, network resource utilization, cost and energy efficiency is conducted.

1.3 Thesis Organisation

This thesis is structured as follows: Chapter 2 reviews existing literature on the different aspects which are vital for the design and implementation of flexible optical metro/core and DC networks. Existing literature on the hardware technologies, optical metro/core and DC network

infrastructure, software technologies which enable network programmability and control, resource allocation, and flexibility as a measurable KPI are reviewed and discussed. The literature review reported in this Chapter has been used in publications [A], [B], [C], [D] and [E].

Chapter 3 reports on the flexibility measurement methodology utilized in this thesis. Following this, the derivation of the proposed flexibility measurement models for different configurations of WSS and SSS are reported. The design trade-off between flexibility and design features of the studied switches are discussed and highlighted. The contributions reported in this Chapter have been used in publications [C] and [E].

Chapter 4 reports on the derivation of the proposed flexibility measurement models for different configurations of BVTs and optical transmission and switching systems. The design trade-offs between flexibility, other KPIs and design features of the different BVT configurations and optical systems are discussed and highlighted. The contributions reported in this Chapter has been used in publications [C], [D] and [E].

Chapter 5 describes features of a reconfigurable hybrid disaggregated DC architecture and two conventional hybrid disaggregated DC architectures. Following this, several proposed network strategies and algorithms to select EPS/OCS services, create custom network topologies, and allocate network resources to build VMs across the various hybrid disaggregated architectures are reported. The contributions reported in this Chapter have been used in publications [A] and [B].

Chapter 6 presents a comprehensive network performance analysis of the different proposed networking strategies across the various hybrid disaggregated DC architectures. Two simulation scenarios are presented. The first simulation scenario is a performance evaluation of the dRedBox architecture and two different conventional hybrid disaggregated DC architectures. The second simulation scenario is a performance evaluation of dRedBox with different arrangement of disaggregated resource pools. The contributions reported in this Chapter have been used in publications [A] and [B].

Chapter 7 presents the conclusion and future work of this thesis.

1.4 List of Publications

1.4.1 Publications Directly Related to this Thesis

[A] **Adaranijo Peters**, George Oikonomou and Georgios Zervas, “In compute/memory dynamic packet/circuit switch placement for optically disaggregated Data Centres,” *Journal of Optical Communication and Networking*, vol. 10, no. 7, pp. 164-178, July 2018.

[B] **Adaranijo Peters** and Georgios Zervas, "Network synthesis of a topology reconfigurable disaggregated rack scale datacentre for multi-tenancy," *in Optical Fiber Communications Conference and Exhibition (OFC)*, Los Angeles, CA, 2017

[C] **Adaranijo Peters**, Emilio Hugues-Salas, Matthias Gunkel and Georgios Zervas, “Key performance indicators for elastic optical transponders and ROADMs: The role of flexibility,” *Optical Switching and Networking,* Volume 25, Pages 1-12, 2017, (**Fabio Neri Best Paper Award**).

[D] **Adaranijo Peters**, Emilio Hugues-Salas, Georgios Zervas and Dimitra Simeonidou, "Design of elastic optical nodes based on subsystem flexibility measurement and other figures of merit," *in European Conference on Optical Communication (ECOC)*, Valencia, 2015.

[E] **Adaranijo Peters**, Emilio Hugues-Salas, Georgios Zervas and Dimitra Simeonidou, "Measuring flexibility and design trade-offs of N × M SSS-based ROADMs and BVTs," *in Optical Fiber Communications Conference and Exhibition (OFC)*, Los Angeles, CA, 2015.

1.4.2 Other Publications

[F] Hui Yuan, Arsalan Saljoghei, **Adaranijo Peter**s and Georgios Zervas, "Disaggregated Optical Data Center in a Box Network using Parallel OCS Topologies," *in Optical Fiber Communication Conference (OFC)*, San Diego, California, USA, 2018

[G] Hui Yuan, Arsalan Saljoghei, **Adaranijo Peters** and Georgios Zervas, "Comparison of SDM-WDM based Data Center Networks with equal/unequal core pitch Multi-Core Fibers," *in Optical Fiber Communication Conference (OFC)*, San Diego, California, USA, 2018.

[H] Heng Liu, **Adaranijo Peters**, Miquel Garrich, and Georgios Zervas, “OSNR Aware Composition of Open and Disaggregated Optical Node and Network” *Journal of Optical Communication and Networking*. Vol. 9, pp. 844-854, 2017.

[I] Yifan Liu, Hui Yuan, **Adaranijo Peters** and Georgios Zervas, "Comparison of SDM and WDM on Direct and Indirect Optical Data Center Networks," *in European Conference on Optical Communication (ECOC)*, Dusseldorf, Germany, 2016.

[J] Shuangyi Yan, Emilio Hugues-Salas, Victor J. F. Rancano, Yi Shu, George M. Saridis, Bijan Rahimzadeh Rofoee, Yan Yan, **Adaranijo Peters**, Saurabh Jain, Timothy May-Smith, Periklis Petropoulos, David J. Richardson, Georgios Zervas, and Dimitra Simeonidou, "Archon: A Function Programmable Optical Interconnect Architecture for Transparent Intra and Inter Data Center SDM/TDM/WDM Networking," *Journal of Lightwave Technology*, vol. 33, no. 8, pp. 1586-1595, 2015.

[K] Shuangyi Yan, Emilio Hugues-Salas, Victor J. F. Rancano, Yi Shu, George M. Saridis, Bijan Rahimzadeh Rofoee, Yan Yan, **Adaranijo Peters**, Saurabh Jain, Timothy May-Smith, Periklis Petropoulos, David J Richardson, Georgios Zervas and Dimitra Simeonidou., "First demonstration of all-optical programmable SDM/TDM intra data centre and WDM inter-DCN communication," *in European Conference on Optical Communication (ECOC)*, Cannes, 2014.

2 Literature Review

2.1 Introduction

This chapter introduces and critically reviews the different aspects of delivering/improving flexibility in optical metro/core and DC networks. The different aspects reviewed are: hardware technologies, network infrastructures, software technologies which enable network programmability and control, resource allocation policies, and the study of flexibility as a measurable KPI. First, the meaning of flexibility is discussed. Next, the features of enabling hardware technologies for the design of ROADMs and EONs are discussed. Following this, network infrastructures for the design and implementation of optical metro/core networks are presented. Particularly, different approaches for ROADM design are reviewed, and the characteristics and benefits of EONs are discussed. Furthermore, disaggregation of DC server resources and different DC architectures which are essential parts of DC infrastructures are reviewed. Particularly, the benefits and challenges of disaggregation of DC server resources are presented, and the different approaches which have addressed the challenges and demonstrated the benefits of disaggregation of DC server resources are reviewed. Also, the features, benefits and shortcomings of different DC architectures are evaluated. Next, software technologies such as NFV and SDN which are key enablers of network programmability and control are introduced. The different approaches which have been proposed to solve the resource allocation problem associated with the optimum placement of virtual network functions in DCs are reviewed. Finally, literature on flexibility as a measurable KPI in the context of optical networks, SDN, and NFV are reviewed.

2.2 What is Flexibility?

The term “flexibility” can be defined in a wide variety of contexts and sectors, hence, the definition and interpretation of flexibility remains ambiguous and subjective to a particular context or sector. Also, other terms such as “programmability”, “reconfigurability” and “elasticity” are synonymous to the term “flexibility” and have been used interchangeably by the engineering community. In the context of optical networks, N. Amaya, et al [4] defined flexibility as the ability of a system to adapt to change. The authors further defined flexibility based on the design features and functionality of an optical network element. For example, an optical component which has the ability to route wavelength channels from an input point to

an output point through different paths has routing flexibility, an optical component which has the ability to rearrange its design structure to build different configurations or architectures has architectural flexibility, an optical component which has the ability to establish a connection between input ports and output ports in multiple dimensions (space and spectrum) has switching flexibility, and an optical subsystem which has the ability to support variable bitrates and select and route the wavelength channels to multiple destination has channel and routing flexibility.

As previously mentioned in the introduction chapter, there has been an extensive amount of research conducted in different essential aspects (i.e., hardware technologies, network infrastructure, software technologies, and resource allocation) that deliver flexibility in optical metro/core and DC networks. However, these studies evaluated flexibility in respect to other measurable KPIs and not considering flexibility as a measurable KPI. For instance, a recent comprehensive tutorial on flexible optical networks in [25] discussed the different building blocks of flexible optical networks but did not review or discuss any work related with the study of flexibility as a measurable KPI. In this regard, the study presented in this thesis touches on the different aspects of delivering flexibility with the inclusion of the study of flexibility as a measurable KPI.

2.3 Hardware Technologies

2.3.1 Wavelength and Spectrum Selective Switch

The WSS is a fundamental building block of ROADMs. A WSS can select and switch wavelength channels between multiple input and output ports, perform power equalization and attenuation of wavelength channels, and support bidirectional transmission through software control [26][27]. A WSS is made up of free space optics integrated with switching technologies such as liquid crystal on silicon [28]–[30] and micro electromechanical mirrors (MEMs) [31]–[33]. Once a light path signal is injected into the input port of a WSS, the light path is directed onto a diffractive grating. The diffractive grating separates the light path into individual wavelength channels and directs the separated wavelength channels onto a switching element. The switching element then redirects individual wavelength channels to the desired output ports through the diffractive grating [34].

Over the past few years, researchers and engineers have developed and introduced the $1 \times M$

WSS which has played a crucial role in developing ROADMs currently deployed in existing optical networks. However, the port dimension which limits the switching functionality of the WSS from one input port to M output ports or from N input ports to one output port limits network flexibility and connectivity. This limitation drove the need for a WSS with multiple input and output ports, i.e., $N \times M$ WSS. The $N \times M$ WSS can be implemented by coupling an $N \times 1$ WSS with a $1 \times M$ WSS in series or by coupling N number of $1 \times M$ WSSs with M number of $N \times 1$ WSSs in parallel [8][26]. However, these techniques of implementing the $N \times M$ WSS have disadvantages. The $N \times M$ WSS design with WSSs coupled in series has wavelength contention because the two WSSs are connected through a single fibre link, and the $N \times M$ WSS design with WSSs coupled in parallel is not cost efficient because a high number of $1 \times M$ and $N \times 1$ WSSs are required to implement an $N \times M$ WSS with a large port-dimension. Thus, the need for an $N \times M$ WSS as a single component without wavelength contention is of crucial importance. A standalone $N \times M$ WSS without contention offers benefits such as supporting multi-flow applications from BVTs [35] and providing contentionless features in ROADMs [8].

The ITU-T G.694.1 has recommended a channel spacing of 12.5 GHz for flexible grid networks [9], however, the filtering functionality of the WSS is limited to fixed grid networks with a rigid channel spacing (50GHz or 100 GHz) and cannot meet this requirement. This challenge has been overcome with the introduction of SSS. The SSS provides a finer spectral granularity which supports flexible grid networks. Thus, an SSS can route and switch spectral slots of different widths, and flexible-grid ROADMs are realized by upgrading WSSs to SSSs [4][36][37]. Though a standalone $N \times M$ WSS/SSS without contention is highly desirable because of the numerous benefits, this design is not technically mature. The commercially available $N \times M$ WSS/SSS can be seen in [38][39], these designs provide I/O port dimension reconfigurability but suffer from internal contention.

2.3.2 Multicast Switch

The MCS is an alternative essential optical component for ROADM design. The MCS is fabricated with silica-based planar lightwave circuit technology and is implemented using a multicast switch design consisting of N number of $1 \times M$ optical splitters connected to M number of $N \times 1$ optical switches [40][41]. The MCS is reliable and supports large scale production. However, due to the presence of optical splitters, as the port dimension of the MCS

increases, the optical loss also increases. In addition, unlike the WSS which has filtering functions to select and block wavelength channels, an MCS does not have this functionality, thus, M number of tuneable filters have to be externally connected to optical switches to achieve wavelength filtering [40].

2.3.3 Flexible Optical Transmitters

Flexible optical transmitters are important for implementing ROADMs and delivering efficient optical network operations. Different transmission requirements and levels of network performance can be achieved by varying programmable transmission features such as number of optical carriers, modulation formats, symbol rates and tuneable wavelength channels. Numerous studies [42]–[44] have demonstrated single carrier software defined optical transmitters with high performance digital to analogue converters (DAC) and digital signal processing (DSP) techniques where different modulation formats and multiple bitrates were achieved. Furthermore, the introduction of multicarrier transmitters technologies such as electrical and optical orthogonal frequency division multiplexing [45]–[47], optical arbitrary waveform generator [48] and Nyquist wavelength division multiplexing [49] have further increased the capacity, flexibility and connectivity of optical transmitters. It is worth mentioning that various multicarrier technologies used for developing BVTs have been discussed in [10]. Also, in order for fixed grid optical transmitters to meet the requirement of flexible grid networks, they have to be equipped with lasers which support finer spectral tuning granularities [50].

Sliceable bandwidth variable transponders (SBVTs) which are also referred to as multi-flow optical transponders have been reported in [10][51][52] as an innovative technology which significantly increases flexibility, connectivity and efficiency of optical transmission. In an SBVT, different combination of optical paths with equal or variable bandwidth form supper channels or multiple optical flows to serve a single or multiple traffic demands. Different designs of SBVTs have been proposed and experimentally demonstrated in [53][54]. Also in [55], a cost evaluation of SBVT in a network scenario showed that SBVT decrease transponders cost.

2.4 Optical Metro/Core Network Infrastructure

2.4.1 Reconfigurable Add/Drop Multiplexer

ROADMs are fundamental building blocks of flexible optical networks. ROADMs provide routing and switching functions of wavelength channels through remote software control [8]. A ROADM node architecture consists of an express cross-connect structure and add/drop cross-connect bank. The express structure provides non-blocking interconnection between different node degrees, and the add/drop bank provides interconnection between the node degrees and transmitters/receivers of the resident ROADM node [5][8]. ROADMs with colourless, directionless and contentionless (CDC) functionality further improve the overall flexibility and performance of optical networks by providing improved connectivity and efficient use of wavelength resources [5][8][56]. The colourless feature is when there are no restrictions in the selection of any wavelength from any port on the add/drop bank. The directionless feature is when there are no directional routing restrictions of wavelength channels to/from any port on the add/drop bank. The contentionless feature is when there are no wavelength conflicts in the add/drop bank (i.e., multiple copies of the same colour of wavelength channel can be simultaneously added from the add/drop bank to any node degree and dropped from any node degree to the add/drop bank).

Numerous studies have presented and proposed add/drop banks with CDC features using several optical technologies which provide different levels of performance. The authors in [57] proposed a CDC ROADM with an add/drop bank which is implemented in two parts. The first part consists of multiplexers interconnected to bulky optical switches and the second part consist of optical splitters. The study in [58] presented an extended version of the previously discussed add/drop bank in [57], the first part of the add/drop bank is implemented with WSSs interconnected to optical splitters while the second part consists of optical splitters interconnected to smaller optical switches. The study compared the two architectures and reported that the add/drop bank with WSSs and smaller optical switches is better than the add/drop bank with large optical switches in terms of modularity, but has a disadvantage of an increased optical loss for both the add and drop paths. The disadvantages of the two add/drop bank designs in [57][58] are the large sizes of the optical switches, and questionable scalability of the add/drop bank as there is a possible increase in optical loss if the port count of the optical splitters are increased. Furthermore, studies in [40][41] presented a CDC ROADM with an add/drop bank which is implemented with an MCS (i.e., a silica-based planar lightwave circuit

transponder aggregator). The authors also presented a fabricated prototype of the MCS and emphasized of the advantages of the proposed MCS in size (i.e., a compact design structure) in comparison to the add/drop bank structure in [57], which is implemented using a bulky optical switch. The authors in [59] proposed a CDC ROADM architecture in which the express cross-connect structure is implemented with 1×43 WSSs, and the add/drop bank with an MCS. The CDC ROADM was experimentally demonstrated with a DP-QPSK transmission system without any significant penalties. As previously mentioned in section 2.3.2, the major disadvantage of the MCS is that as the port dimension increases, there is an increase in optical loss due to the presence of optical splitters. Thus, optical loss is an issue for scaling add/drop banks implemented with MCSs.

The add/drop bank of CDC ROADM can also be implemented with different design configurations of WSS [8][26][35][59]: (i) interconnection of an $N \times 1$ WSS with a $1 \times M$ WSS in series, (ii) interconnection of N number of $1 \times M$ WSSs with M number of $N \times 1$ WSSs in parallel, and (iii) a standalone contentionless $N \times M$ WSS. In the add/drop bank implemented by interconnecting an $N \times 1$ WSS with a $1 \times M$ WSS in series, because there is internal wavelength contention present in a single module, a sizeable number of modules are required to eliminate wavelength contention within the add/drop bank. This leads to an increase in cost. The add/drop bank design implemented by the interconnecting N number of $1 \times M$ WSSs with M number of $N \times 1$ WSSs in parallel can deliver CDC functionality, however, this design introduces design complexity and cost because a large number of switches are required to implement an add/drop bank with a large port count. The add/drop bank implemented with a standalone contentionless $N \times$ M WSS overcomes the drawbacks of the previously discussed WSS-based add/drop banks, this is because it can provide CDC features from a single device and has lower optical loss in both the add/drop direction (i.e., the optical signal travels through only one WSS in the add/drop bank). However, the limitation of this design of WSS is that it is not yet technically mature.

There have been several studies which have evaluated the performance of different add/drop bank architectures and design features. The authors in [56] presented a study which evaluated the effect of add/drop bank CDC features on the performance of optical networks. Results from analytical models and simulations showed that in terms of blocking probability, an add/drop bank with CDC features performs better than an add/drop bank with only colourless and directional features, and an add/drop bank which has colour, direction, contention restrictions.

The study in [60] presented a performance evaluation of various ROADM node architectures in terms cost, optical loss and size of equipment. Another study in [61] presented a comparative evaluation of an MCS based add/drop bank and an optical cross-connect WSS based add/drop in terms of OPEX and CAPEX.

The flexibility of ROADM architecture has further been increased with the introduction of the architecture on demand (AoD) concept. The concept of AoD which was first introduced in [62], consist of an optical backplane (implemented using a high port count MEMs optical switch) which interconnects optical modules such as WSS, multiplexers, demultiplexers, couplers and optical amplifiers. Unlike traditional ROADM architectures where the interconnection between components are hardwired, in an AoD node architecture, the interconnection between modules are not hardwired. Therefore, different architectures can be setup to handle different traffic requirements and support various network functions such as fibre switching, spectrum switching and packet switching without overuse of resources. For instance, assuming all wavelength channels from node degree 1 are to be passed to node degree 2, in a broadcast and select (BS) ROADM architecture, all wavelength channels will be routed through an optical splitter and WSS, whereas in an AoD node architecture, only a single fibre switch interconnection from node degree 1 to node degree 2 is required. However, even with the promising benefits, the AoD concept still has some challenges [62]. Firstly, the introduction of this amount of flexibility requires efficient resource allocation algorithms for construction of architecture while considering dynamic network requirements, cost and energy efficiency. Secondly, deploying an AoD node in a network scenario with dynamic traffic trends is challenging because the switching time required to setup an optical cross-connection on the optical backplane can hinder hitless routing and switching operations.

Since the introduction of the AoD concept, various studies have proposed strategies and algorithms to address the challenges and demonstrate the potential benefits. The authors in [63] proposed a synthesis algorithm to address challenges associated with the number of network components utilised when building architectures in an AoD node. Results from the study showed that in comparison to different traditional hardwired ROADM architectures, an AOD node can decrease the number of hardware components utilised. A similar study in [64] proposed an extended synthesis algorithm addressing different types of switching traffic requests and the selection of network components when building AoD node architectures. Results showed that the AoD can achieve better performance than traditional ROADMs in

terms of power consumption and cost. The authors in [65] proposed a strategy and an algorithm to decrease power consumption in AoD nodes by avoiding the utilisation of unnecessary amplifiers and turning them into sleep mode. The results from the simulation study showed that networks equipped with AoD nodes can achieve a reduction in power consumption and improve network availability in comparison to networks equipped with traditional ROADMs. Despite the benefits of the AoD concept demonstrated in [63]–[65], none of these studies considered a dynamic traffic scenario and the negative impact of the switching time introduced by the optical back plane of the AoD. Another study in [66] proposed ILP formulations and heuristic algorithms to address network planning problems and deliver cost-efficient optical networks with AoD nodes in static and dynamic network scenarios. Results from the study showed that optical networks equipped with AoD nodes achieved switching module savings and lower power consumption in comparison to optical networks equipped with traditional hardwired ROADMs. However, in order to reduce network blocking in the dynamic network scenario, the location for the placement of optical components at different AoD nodes was implemented using approximations from a set of random static traffic. This approach does not fully represent a dynamic network scenario where network requests are not known in advance. The feasibility of the AoD concept has been enhanced with experimental demonstration in [4][67][68].

This section reviewed numerous studies which have proposed different designs and approaches to deliver/improve flexibility in ROADMs and evaluated different ROADM architectures in terms of energy efficiency, scalability, cost and optical loss. However, the study of flexibility as a measurable KPI for the design of ROADMs is minimal, only studies in [4][5] have investigated flexibility as a measurable KPI for the design of ROADMs. These studies are reviewed and evaluated in section 2.7.

2.4.2 Elastic Optical Networks

The spectrum of traditional wavelength division multiplexing (WDM) networks is divided into fixed channel spacings of 50GHz or 100 GHz, this rigid framework leads to wastage of bandwidth and inefficient utilisation of the spectrum [10][36]. EONs have been introduced as a solution to overcome the drawback of traditional WDM networks and provide greater flexibility for optical transmission. EONs support flexible utilisation of the spectrum such that a suitable size of spectral slot is allocated to an optical signal to prevent spectrum wastage [10][11][36]. The realization of EONs has been made possible with the development of

hardware technologies like the BVTs and SSSs (discussed in section 2.3). In addition, several studies [69]–[73] have evaluated and demonstrated the benefits of EONs over traditional WDM networks in terms of KPIs such as energy efficiency, spectrum utilisation and cost.

Despite the promising benefits of EONs, there are numerous challenges associated with delivering efficient and optimum network performance. In traditional WDM networks, a key challenge for delivering optimum performance when serving a traffic demand is finding a suitable network path and allocating a wavelength channel while satisfying the wavelength continuity constraint. This problem is commonly known as the routing and wavelength assignment (RWA) problem [74]. In EONs, apart from selecting a suitable network path, spectrum contiguity and spectrum continuity constraints must also be satisfied in order to successfully establish a lightpath. This problem in EONs is known as the routing and spectrum allocation (RSA) problem [75][76]. The RSA problem can also be extended to the joint optimization other transmission features (e.g. modulation formats, number of carriers or symbol rates) to improve network performance and solve spectrum fragmentation problems which occurs as a result of stranded bandwidth introduced by dynamic arrival and release of traffic demands [75].

An extensive amount of studies have proposed ILP and heuristics algorithms to solve the RSA problems and other challenges associated with EONs such as spectral fragmentation. The study in [77] proposed RSA algorithms to improve spectrum utilization and fragmentation in EONs with mixed line rates. The proposed algorithms were evaluated in comparison to existing RSA algorithms in a dynamic network scenario. Results from the study showed that the proposed RSA algorithms can minimize blocking probability and spectrum fragmentation. Other studies have proposed RSA algorithms in EONs considering additional transmission constraints such as modulation formats [70][78][79] and transponders [80]. In addition, the authors in [75] presented a comprehensive tutorial and review on RSA methodologies for EONs. The study discussed different strategies for solving RSA problems and spectrum fragmentation in dynamic and static network scenarios.

2.5 Data Centre Network Infrastructure

2.5.1 Disaggregation of Data Centre Server Resources

Traditional server-centric DCs consist of a rigid resource framework where each server is

composed of a fixed amount of CPU, memory, storage and network resources which are tightly packed and coupled onto a single mother board [12]. This server design has limitations in resource utilisation and modularity due to the disparity between different types of DC workloads and the static server resource configurations [12][81]. Disaggregation of DC server resources into standalone resource pools (i.e., CPU, memory, storage, and network) has been proposed as a solution to address the limitations of traditional server-centric DCs, and offer benefits such as an increase in server resource utilization, modularity and flexibility [12]. However, the authors in [12] pointed out the DC communication network design as a key challenge which must be explored to reap the benefits of server resource disaggregation. This is because communication between resources (e.g. CPU to memory) which has traditionally been restricted within a server blade with compactly integrated resources will now be distributed throughout the DC network. Another key challenge for the realization of disaggregation of DCs is IT and the network resource allocation. This is because a wide variety of factors such as resource availability, resource capacity, location of the different resources types across the DC network (i.e., how CPU and memory resources are arranged) and the network fabrics will have to be considered when deploying resources to build VMs in disaggregated DCs.

Recent Studies have proposed algorithms for IT and network resource allocation, and carried out simulations studies to evaluate the performance of disaggregated DCs [21]–[24]. A. Pages *et al.* [21] presented an integer linear programming and heuristic based model to minimize the computing resources required for the deployment of virtual DC traffic requests on disaggregated DC resources in a static network scenario. The resource utilization of a disaggregated DC was evaluated and compared with a server-centric DC with the same resource pool. Results from the study showed that disaggregated DCs offer better resource utilization than server-centric DCs. Also in [22], results from a mixed-integer linear programming and heuristic model showed that disaggregated DCs offer better power savings than server-centric DCs for different types of VMs. The authors in [23] proposed and evaluated several algorithms for dynamic IT and network resource allocation for dRedBox architecture. The results from the study demonstrated that the dRedBox architecture can achieve better resource utilisation than server-centric DCs. Another study in [24] proposed a scheduling algorithm with the objective of maximizing CPU resource utilization without performance degradation when deploying VMs in a disaggregated DC. The study evaluated the performance of the proposed algorithm in comparison to existing algorithms. Results from the simulation

study showed that the proposed algorithm outperformed the existing algorithms in terms of resource utilization and energy efficiency. The existing research work on resource allocation in disaggregated DC is minimal and various challenges associated with the allocation of IT and network resources have not yet been studied. For instance, none of these studies have proposed strategies and algorithms to solve challenges associated with the selection and allocation of EPS/OCS network services to build VMs on hardware programmable disaggregated resources or carried out a comprehensive performance evaluation of their proposed algorithms across different disaggregated DC architectures. Therefore, in order for the concept of server resource disaggregation in DCs to materialize, continuous studies and innovations through the development of efficient resource allocation strategies and algorithms, and the comprehensive performance evaluation of the proposed algorithms across different DC architectures are required.

The concept of disaggregated DC resources have been experimentally demonstrated in [82]–[86]. P. Gao *et al.* [82] investigated network bandwidth and network latency performance requirements for DC disaggregation to avert application performance degradation. The study showed that for certain applications, 20-40 Gb/s for remote memory access can achieve minimal (under 10%) application performance degradation. In addition, the authors in [83] experimentally demonstrated communication between Multi-Processor System on Chip (MPSoC) hardware with multiple cores and remote access memory over 10 Gb/s lanes. The authors reported that a bandwidth of 582 Mib/s (~ 5 Gb/s) can be realized between a single CPU core and remote access memory. Also, the authors in [84] experimentally demonstrated remote memory access with 10 Gb/s links and up to 68% sustained memory bandwidth.

2.5.2 Data Centre Architectures

Pure electrical DCs such as Fat tree [87] which consist of multiple layers of electronic packet switches connected in a tree like approach have experienced shortcomings due to high power consumption, cost intensive-scalability and fixed switching capacity. To overcome this, there has been a shift in focus from conventional electronic-based DCs to DCs which have incorporated either electronic/optical technologies or all-optical technologies. This is because of the significant benefits that optical technology provides (i.e., low latency, energy efficiency and high data rates).

Electrical/optical switching DC architectures

The authors in [14] proposed HELIOS, a hybrid DC architecture which incorporates electronic/optical switching fabrics. HELIOS has a 2-tier network level of core and pod switches. Each pod (ToR) switch houses several hosts and are embedded with 10G transceivers. The layout of the core switches are a blend of electronic packet switches and MEMs optical switches. Half the transceivers on each pod are connected to electronic packet switches while the remaining transceivers are connected to a MEMs optical switch through multiplexers. The electrical network handles bursty inter-pod communication while the optical network handles slowly varying inter-pod communication. The control loop of HELIOS architecture is made up of a topology manager, a pod switch manager and a circuit switch manager. The topology manager is the centre of the HELIOS architecture and its functions include examination of varying communication patterns, computation of new topologies and traffic demand estimation. The topology manager also interfaces with the circuit switch manger and pod switch manager which are run on the circuit switch and pod switches, respectively. The authors also carried out a simulation study by comparing HELIOS to a traditional DC architecture which has a pure electrical packet network. Results showed that HELIOS outperforms the traditional DC with a decrease in cost, power consumption and port count.

Another hybrid DC architecture called c-Through has been proposed in [15]. c-Through consists of a multi-tier electrical packet network with Ethernet switches connected in a tree-like manner and an optical network which consist of a MEMs optical circuit. Both the electrical network and optical network are connected to the ToR switches. In the optical network, each ToR can only be interconnected to another ToR at a time. The differences between HELIOS and c-Through architectures are: (i) The electrical network of HELIOS is a 2-tier network while the electrical network of c-Through is a 3-tier network. This could potentially lead to an increase in power consumption in c-Through because inter-ToR communication in the electrical network may transverse more one network hop. However, in HELIOS, inter-ToR communication in the electrical network is limited to just one network hop. (ii) In the optical network, c-Through only allows each ToR to communicate with another ToR at a time. This point to point communication restriction does not occur in HELIOS. Hybrid electrical/optical DCs can deliver significant benefits over pure electrical DCs in terms of energy efficiency and cost. However, the presence of power hungry electronic switches at the core switching layer still presents a limitation in terms of energy efficiency.

All-optical switching DC architectures

The authors in [88] proposed an optical switching DC architecture called OSA (previously presented as PROTEUS in [89]). In OSA, each ToR (electrical switch) is connected to a MEMs optical switch through an optical module consisting of a WSS, a MUX, a DEMUX, a coupler and optical circulators. The outbound traffic from the ToR is transmitted through an interconnection of a MUX, a WSS and circulators, while inbound traffic to the ToR is transmitted through an interconnection of circulators, a coupler and a DEMUX. Each ToR is interconnected to only *k* other ToRs, and communication to other ToRs that are not directly connected is achieved by multi-hop chaining through the MEMs optical switch and ToR switches. This architecture offers efficient port utilization due to the use of optical circulators, flexibility in link capacity through WDM, and flexibility in topology computation through the reconfigurability of MEMs optical switch. The authors evaluated OSA in terms of network bisection bandwidth. The results presented showed that the OSA can achieve high bisection bandwidth of about 60% -100% of non-blocking networks under different traffic patterns.

Another optical DC architecture called WaveCube has been proposed in [90]. The composition of the WaveCube architecture is similar to OSA, the major difference is that there is no MEMs optical switch in WaveCube. ToR to ToR connectivity is achieved by direct interconnections through the optical modules. WaveCube achieves fault-tolerance since a single point of failure is eliminated by the removal of the MEMs optical switch. Since there is no optical back plane to deliver a dynamic topology, the authors mitigate this disadvantage by proposing an optimum solution for dynamic link bandwidth scheduling and multi-pathing routing. A major drawback of OSA and WaveCube is the possible increase in latency due to the multi-hop chaining of bursty traffic flows.

The study in [91] proposed a Petabit optical switch architecture for DCs which employs arrayed waveguide grating routers (AWGRs) and tunable wavelength converters (TWCs). The optical switch is implemented using a 3-stage switching network where the AWGR is the fundamental switching device at each stage. The TWCs are attached to the input of the AWGRs at the second and third stages to perform wavelength conversion for dynamic configuration. Each ToR switch is interconnected to the optical switch through WDM (MUXs and DEMUXs). In addition, none of the three switching stages needs buffers and fibre delays lines because

buffering occurs at the line cards which are connected to the input of the switch fabric. This design feature reduces latency and system complexity. The authors in [19][92] proposed a DC architecture called HOSA. HOSA is a 2-tier network level architecture. ToR switches which are electronic switches (first-level) are interconnected to slow optical and fast optical switches (second-level). The slow optical switches are based on MEMs technology while the fast-optical switches are based on AWGR or semiconductor optical amplifier technologies. Furthermore, the control of the DC is achieved via a centralized controller which carries out scheduling, routing and switching configuration tasks.

A flat optical DC architecture based on optical circuit and optical packet switching (OPS) called LIGHTNESS has been proposed in [93]. The data plane of LIGTHNESS is implemented based on the AoD concept. An optical backplane (i.e., a high port-dimension MEMs based optical switch) interconnects ToR switches, optical packet switch fabric, MUXs and DEMUXs. OCS is provided by the MEMs based optical switch to serve high bandwidth long-live traffic flows while OPS is provided by the optical packet switch fabric to serves short-lived traffic flows. The control plane of LIGHTNESS is achieved through SDN, which provides network control and management of the DC infrastructure. The authors in [18] carried out a simulation study and proposed a VM placement strategy to enable multi tenancy in LIGHTNESS and compared it with a pure optical switching DC. Results showed that LIGHTNESS achieved a higher VM acceptance ratio than the pure optical switching DC under the same traffic and network configurations. Another study in [17] proposed an ILP model and heuristic algorithms for minimizing the number of transceivers used when deploying virtual slices in LIGHTNESS and compared it with a pure OCS DC architecture. An evaluation of the results showed that LIGHTNESS performs better that the pure OCS DC architecture in terms of resource utilisation under the same traffic distribution and network configuration as a result of the statistical multiplexing function of the OPS technology. Despite the promising advantage and flexibility of the LIGHTNESS architecture, the main shortcoming of this architecture is that technical maturity and feasibility of OPS switching technology is still uncertain.

Despite the benefits of the previously discussed electrical/optical and all-optical DC architectures, the collective limitation of these DC architectures is that their network function services are hardwired to specific network and server ports and cannot be reconfigured. This limits the optimization of service offering at the planning and dimensioning phase and in turn the reaction to unpredictable future demands which can lead to a decline in performance. This

limitation has created a need to introduce dynamic network function service programmability in DCs. As previously mentioned in the introduction chapter, the dRedBox architecture which is examined in this thesis overcomes the limitations of traditional server-centric DC architectures. The dRedBox architecture combines server resource disaggregation with software, optical and electronic technologies, and supports deep network function programmability (the dRedBox architecture is described in detail Chapter 5).

2.6 Network Programmability and Control

2.6.1 Network Function Virtualisation

Network hardware devices deployed in traditional networks are proprietary to specific network equipment vendors and manufactured to perform specific network functions [94][95]. This limits network functions service reconfigurability to handle dynamic traffic trends, and introduces unnecessary incurred CAPEX and OPEX because upgrading/changing network services in existing network infrastructure will require the manual installation of additional network equipment by experts. NFV has been proposed as a solution to address these challenges. NFV separates network function from hardware and executes network functions via software on commodity hardware devices (servers, switches and storage) which can be remotely and dynamically deployed to different locations within the network [94][95]. A study in [20] proposed and demonstrated NFV service reconfigurability on an FPGA platform made up of partial reconfigurable regions interconnected through a network of chip. The authors demonstrated network function service reconfiguration on run-time between layer 1 circuit switching service and layer 2 Ethernet packet switch service without service disruption or packet loss. In addition, a programmable hybrid DC with network function service programmability was presented and demonstrated in [96]. This work involved replacing static NICs with FPGA based interfaces on server resources and deploying network service chains to deliver multi-layer networking services.

An important issue associated with NFV is resource allocation for the implementation of virtual network functions (VNFs) [97]. The introduction of network function reconfigurability introduces challenges on how and where VNFs should be deployed in order to achieve optimum network performance. Thus, efficient resource allocation strategies and algorithms must be developed to address this issue. There has been an extensive amount of studies which have proposed ILP formulations and heuristic algorithms for the placement of VNFs across

different types for NVF network environments such as intra/inter DC networks and mobile networks. The authors in [98] proposed a binary integer programming formulation and heuristic algorithm to reduce the number of O/E/O conversions for VNFs placement and NFV service chaining in packet/optical DCs. Results showed that in terms of number of O/E/O conversions, the proposed heuristic algorithm delivers an optimum performance similar to the binary integer programming solution, and a better performance than a first-fit algorithm (i.e., utilizes less O/E/O conversions). The authors in [99] proposed an ILP formulation and heuristic algorithms for VNF service chaining on an optical DC architecture. The VNF service chaining was divided into 3 interrelated issues which includes VNF chaining multiplexing problem, VNF placement problem and RWA problem. These problems were addressed in a sequential manner. Results from a performance evaluation showed the efficiency of the proposed algorithms in terms of implementing VNF service chains on network and IT resources. The authors in [100] proposed an optimization solution (minimum vortex cover and bin packing) and a greedy heuristic algorithm for the placement of security monitoring VNFs between VMs or VMs and external users in a DC network. Results showed that the performance of the proposed greedy algorithm is close to the optimum solution. Another study in [101] proposed an ILP formulation and algorithms to address the network function service chaining placement issue in a multi-layer tree like DC. Results showed that in comparison to conventional best-fit algorithm, the proposed algorithms can minimize bandwidth consumption while delivering similar performance in terms of server utilisation. The authors in [102] proposed a mixed integer programming formulation and heuristic algorithm for VNF placement and antenna scheduling in a hybrid DC. The hybrid DC is composed of ToR, core and aggregate switches. In addition, each ToR switch is equipped with antennas which creates a wireless network. Simulation results showed that the performance of the proposed greedy algorithm is close to the optimum solution. The main drawbacks of the previously discussed studies on VNF placement in DCs [98]–[102] are that the studies did not consider dynamic network scenarios for VNF placement, and validate the performance of the proposed algorithms across different DC architectures. In addition, work on resource allocation for VNFs placement in disaggregated DCs is still unexplored.

In the case of VNF placement in inter-DC networks, a joint allocation of wavelength/spectrum resources from optical metro/core networks which interconnects DCs and IT resources across the DCs are required for the placement and service chaining of VNFs [103]. The authors in [103] proposed an ILP formulation and heuristic algorithm for optimum placement and service

chaining of VNFs in inter DC EONs. The objective was to solve the issues associated with the joint allocation of spectrum and IT resources in a static network scenario. The proposed algorithms were compared against benchmark algorithms, and results from the study showed that the proposed ILP and algorithms can efficiently decrease the maximum utilised frequency slot and reutilize VNFs. Other studies in literature [104]–[106] have also proposed ILP formulations and heuristic algorithms to solve the joint allocation problem of IT and spectrum resources associated with VNF placement in inter DC EONs.

2.6.2 Software Defined Networking

The data and control plane of traditional networks are placed within the proprietary network hardware devices, therefore, manual intervention is required in order to change/upgrade management and control policies of the network [107][108]. This limits operational flexibility and network scalability, and leads to disruption of network services. Software defined networking (SDN) has been proposed a solution to eradicate this issue. In SDN, the control plane functionality is separated from the data plane of network devices and is placed in a central SDN controller which manages the entire network and enables network programmability [107][108]. The introduction of SDN has opened up pathways to achieve improved network management and configuration, enhance network performance and optimization, and simplify implementation and testing of novel ideas [109]. However, there are some challenges relating to issues such as scalability, performance and security that have to be addressed to fully realise its full potential [110]. The reference model for SDN comprises of three layers [107][109][111]: infrastructure, control and application layer. The infrastructure layer is made up of interconnected switching devices (e.g. routers) which process and forward packets based on the network policy of the controller. The control layer is a link between the infrastructure and application layer, it comprises of a controller which maintains and keeps records of the state of network and provides instructions to hardware devices using protocols. The application layer consists of network service applications for users.

2.7 Flexibility as a Measurable Key Performance Indicator

Measurable KPIs aid network operators and engineers to determine optical network behaviour and understand the relationship between design features and performance output, which in turn leads to efficient design and management of network infrastructure and operations. There are numerous studies on flexibility as a measurable KPI in the field of manufacturing [3][112]–

[116]. These studies reviewed and discussed existing literature on manufacturing flexibility, presented various definitions for manufacturing flexibility in different contexts such as machine flexibility, expansion flexibility, process flexibility etc., and discussed different approaches to measure flexibility. In particular, studies in [115][116] have proposed entropy as an approach for measuring flexibility in manufacturing systems.

In the context of SDN and NFV, studies in [6][7] have discussed the importance of flexibility as a measurable KPI for network design and proposed methods to quantify flexibility. The study in [6] proposed an approach for measuring the flexibility of mobile core networks with SDN and NFV technologies. The study evaluated three design choices (SDN only, NFV only, and a combination of SDN and NFV) across logically centralized and distributed DC infrastructure. The flexibility was measured based on a weighted feasibility solution for function placement under control and latency constraints. Results from the study showed that a combination of SDN and NFV provides the greatest flexibility for centralized DC infrastructure scenario. For the distributed DC infrastructure, results from the study showed that the flexibility of all three design choices are approximately equal. The study in [7] proposed a flexibility measurement approach for SDN networks and evaluated design trade-offs between flexibility, cost and time. The authors defined the flexibility as a measure of the portion of requests that can be sustained from a known set of requests within a time limit. The network flexibility was evaluated in two different cases. The first case evaluated network flexibility with respect to SDN controller placement, and the second case evaluated network flexibility with respect to failure recovery. In the first case, the flexibility is expressed as the ability of the SDN control plane to adjust to varying flows of traffic within a migration time limit. An evaluation of four systems with different controller numbers was carried out. Results showed that the flexibility of the different systems increases as the migration time limit increases. This is due to the fact that more reconfigurations can occur within larger values of a migration time limit. Furthermore, at large values of a migration time limit, there is a reduction in the operational cost of the systems as the flexibility increases. The reduction in cost is more evident as the number of controllers increase. In the second case, the flexibility was expressed as the ability of the network to recover from failure within a specified time limit. Three different approaches were evaluated. In the first approach which is called 1+1 protection, resources are duplicated into a separate path-pair and the time for recovery is almost immediate. In the second approach called 1:1 protection, the recovery time is increased despite the fact that resources are reserved in advance, this is because redirection of the traffic through the backup

path are compulsory after failure detection. In the third approach which is called restoration, the recovery time is highest because no resources for protection are duplicated or reserved prior to a failure occurrence. Results from the study showed that the 1+1 protection approach demonstrated the highest cost but with a constant and high level of flexibility irrespective of the recovery time limit, while the restoration approach demonstrated the lowest cost and highest flexibility which is only achieved after a certain value of recovery time limit has been surpassed.

Although these studies provided significant insight into network flexibility as a measurable KPI for network design in the context of SDN and NFV, some key issue were not considered. First, a generic framework and approach for developing flexibility measurements models were not provided. The lack of a generic framework and approach introduces complexity because different approaches for measuring flexibility will need to be developed for the wide variety of networking scenarios. Second, the impact of network devices on network flexibility were not taken into consideration when developing the flexibility measurement approaches.

In the context of optical network hardware and infrastructure, studies in [4][5] have evaluated and discussed flexibility as measurable KPI. The authors in [4] derived and proposed flexibility measurement models based on maximum entropy for key optical components of optical nodes, an optical switching subsystem, and several elastic optical node architectures. The optical components studied include a 1×1 WSS, a $1 \times N$ WSS, a $1 \times N$ SSS, and a $N \times N$ optical switch. The optical subsystem studied was a combination of $1 \times N$ DEMUX and $N \times N$ optical switch. In terms of elastic optical node architecture, the design features and working principle of a BS node architecture, a spectrum routing (SR) node architecture, a switch and select with dynamic functionality (SSDF) node architecture, and an AoD node architecture were discussed and evaluated in terms different types of flexibility, through loss and hardware resource. In the BS node architecture, the ingress and egress nodes are implemented with optical splitters and SSSs, respectively. This architecture provides switching flexibility, no routing flexibility as only a single path between an ingress and egress node degree can be established, and limited functional flexibility (i.e., restricted to only functions provided by SSS such as filtering). In the SR node architecture, both the ingress and egress nodes are implemented with SSSs making it more expensive than the BS architecture. However, it does not have questionable scalability like the BS node architecture where optical losses are determined by the number of node degrees (i.e., port count of optical splitter). This architecture provides switching flexibility, no

routing flexibility and limited functional flexibility. In the SSDF node architecture, the ingress nodes and egress nodes are implemented with optical splitters and SSSs, respectively, and are interconnected via an optical switch. Different modules can be plugged to the optical switch to provide different functions such as packet switching. This architecture provides switching flexibility, routing flexibility as several network paths between ingress and egress node degrees can be established via the optical switch, and functional flexibility as different network function modules can be plugged to the optical switch. In AoD node architecture, different architectures can be constructed to suit different requirements, thus, AoD node architecture provides architectural flexibility. In addition, the AoD node architecture provides switching flexibility, and a higher routing and functional flexibility than the SSDF architecture. Among all the examined optical node architectures, the AoD is the most flexible architecture while the SSDF ranks second. In the case of the BS and SR node architectures, their switching flexibility were measured as a function of number of node degrees. Results from the study showed that the BS architecture has greater switching flexibility than the SR architecture when the number of node degrees exceeds two. This was attributed to the multicasting feature of BS architecture.

The same flexibility measurement approach utilised in [4] has also been utilised in [5] to evaluate the flexibility of different add/drop banks considering the drop direction. In more detail, flexibility measurement models for a WSS based add/drop bank (i.e., $N \times 1$ WSS interconnected to $1 \times M$ coupler), an MCS based add/drop bank, a variable splitter MCS add/drop bank, and AoD add/drop bank were derived and proposed. The flexibility of the different add/drop architectures were measured as a function of node degrees. From the results presented, the MCS based add/drop bank demonstrated a higher flexibility than the WSS based add/drop bank due to avoidance of wavelength contention. The variable splitter MCS add/drop bank demonstrated greater flexibility than the WSS based add/drop bank and MCS based add/drop bank due to additional routing paths. The AoD add/drop bank demonstrated the highest flexibility among the various add/drop banks due to its architectural flexibility.

Despite previous works which provided insights into the benefits of the design/performance evaluation of optical nodes and add/drop banks using flexibility as a measurable KPI, an extensive range of key enabling technologies such as $N \times M$ WSS/SSS, programmable single and multicarrier optical transmitters, and optical transmission and switching subsystems which are vital for ROADM design and delivering flexible optical network have not been studied. In particular, none of these studies considered the different transmission features such as

modulation formats when deriving the flexibility measurements models for evaluating the flexibility of the optical nodes and add/drop banks, or studied the relationship between flexibility and other significant KPIs such as connectivity or capacity.

2.8 Chapter Conclusion

This chapter presented a review of existing literature on different aspects that constitute flexible optical metro/core and DC networks. The aspects reviewed include: hardware technologies, optical metro/core and DC network infrastructure, software technologies for network control and programmability, resource allocation, and the study of flexibility as a measurable KPI. Research gaps, challenges and limitations of existing literature were reported and identified.

Different studies which have proposed and demonstrated various designs and approaches to deliver flexibility in ROADMs and EONs were reviewed. Majority of these studies evaluated the performance and benefits of their various designs and approaches of delivering flexibility in terms of KPIs such as energy efficiency, scalability, blocking probability, cost, and optical loss, but without considering flexibility as a measurable KPI. The study of flexibility as a measurable KPI for network design is a research area that has generally been unexplored by the research community due to the subjective interpretation of flexibility, and the challenges and complexity involved in developing flexibility measurement models. The few existing studies which examined flexibility as a measurable KPI in the context of optical networks, SDN, and NFV were reviewed in this chapter. In the context of optical networks, these studies proposed flexibility measurement models based on maximum entropy for different optical components, elastic optical node architectures and add/drop bank architectures. Also, comparative evaluations of the design/performance of various elastic optical node architectures and add/drop bank architectures in terms of different types of flexibility, design features, and through quantitative measurement of flexibility were reported. Regardless of the progress made in this research area, an extensive range of key enabling technologies such as $N \times M$ WSS/SSS, BVTs and optical transmission and switching subsystems, which are vital for delivering flexibility in ROADMs and EONs have not been studied. Therefore, developing flexibility measurement models for these optical components and subsystems, and evaluating the relationship between flexibility, other KPIs and design features are essential to understanding flexibility as a KPI for network design. This may in turn assist network designers to determine the different quantitative levels of flexibility required to provide cost effective and optimum levels of network performance.

Additionally, existing literature on disaggregation of server resources in DCs, different DC architectures and network function virtualization in DCs were reviewed and discussed. It was reported that disaggregated DCs offer better performance in server resource utilization, flexibility and modularity in comparison to traditional server-centric DCs. However, the communication network design and allocation of resources are key challenges that have to be solved for the concept of disaggregation of DC servers to materialize. Various studies which have proposed IT and network resource allocation algorithms and carried out simulation studies on disaggregated DCs to address the challenges and demonstrate the benefits of DC server resource disaggregation were reviewed. However, these studies did not consider issues associated with the selection and deployment of EPS/OCS services to build VMs on hardware programable disaggregated resources or carry out a detailed performance evaluation of their proposed algorithm across various DC architectures. Furthermore, several DC architectures employing either electrical/optical or all-optical technologies were discussed and reviewed. The benefits and shortcomings of the various architectures were reported. It was established that the collective limitation of these DC architectures is that their network function services are hardwired and dedicated to specific sever and network resources. This limits the ability of these DCs to respond to variable and dynamic networking requirements. Thus, the need for the introduction of network function programmability in DCs. The advantages and challenges of NFV which enables network functions service reconfigurability were discussed. It was reported that a key problem in NFV is resource allocation for the optimum placement of VNF. Various studies which have proposed algorithms for optimum VNF placement in DCs were reviewed. However, it was noted that these studies did not take into account dynamic network scenarios or validate the proposed algorithms across different DC architectures. Also, none of the studies examined VNF placement on disaggregated DC resources.

The combination of server resource disaggregation with electrical/optical technologies and network function programmability has promising potential to eliminate the limitations of traditional server-centric DC architectures. However, there are still critical challenges on how to efficiently deploy and manage these network services to deliver an optimum level of performance when building VMs on programmable disaggregated resources. It is therefore necessary propose and develop novel network strategies and algorithms to address these issues.

3 Wavelength and Spectrum Selective Switch

3.1 Introduction

In this chapter, the approach used to derive and develop flexibility measurement models in this thesis is presented. Next, the flexibility measurement models for the WSS and SSS under the following design conditions are presented: without contention, with contention, and with I/O port dimension reconfigurability. The flexibility of different port-dimension of WSS and SSS are measured, and the relationship and design trade-offs between flexibility and design features are evaluated and discussed. Finally, numerous significant insights from the results and theoretical analysis reported in this chapter are highlighted. Table 3.1 presents the list and definition of all parameters used in the derivation of all flexibility models in this chapter. All flexibility measurements models used to measure the flexibility of optical components in this chapter were implemented in MATLAB.

Table 3.1: Parameter definition for flexibility measurement models of WSS/SSS

Parameter	Definition
N	Number of input ports
M	Number of output ports
P	Total number of ports
W	Number of wavelength channels
k	Spectral granularity factor
kW	Number of spectral slots
a	Number of copies of the same colour of wavelength channel/spectral slot that can be passed from N input ports to M output ports at the same time
x	Maximum number of copies of the same of colour wavelength channel/spectral slot that can be passed from N input ports to M output ports at the same time

3.2 Approach for Developing Flexibility Measurement Models

N. Amaya, et al [4] proposed an approach to measure the flexibility of optical network elements. The authors stated that the flexibility of a system can measured considering the maximum entropy of a system, and the flexibility of a system is obtained by

$$F(S) = log(M) \tag{3.1}$$

where $S = \{s_1, s_2, \ldots, s_m\}$ is the set containing all unique possible states of the system (i.e., assuming all states are equiprobable) and M is the number of elements in the set S. For clarity in the derivation of flexibility measurement models in this thesis, the symbol M is substituted with the symbol T. Thus, the model for flexibility of a system (Equation.3.1) is rewritten as

$$F(S) = log(T) \tag{3.2}$$

Optical subsystems are constructed by combining two or more standalone optical components with different design features. Thus, the design features of the optical components determine the resultant flexibility and functionality of an optical subsystem. According to [4], if an optical subsystem consists of two optical components a and b. Assuming the flexibility of a is F_a and b is F_b. The resultant flexibility F_{ab} of the optical subsystem is

$$F_{ab} \leq F_a + F_b \tag{3.3}$$

The flexibility of the optical subsystem F_{ab} is only equal to the sum of the individual flexibilities of components F_a and F_b when the optical components of the subsystem are disjointed.

3.3 Flexibility Measurement of Wavelength Selective Switch

3.3.1 $N \times M$ WSS without Contention

In an $N \times M$ WSS without contention, copies of the same colour of wavelength channel or copies of different colours of wavelength channels can be switched or blocked from any number of input ports to any number of output ports provided that wavelength collisions are avoided, i.e., switching copies of the same colour of wavelength channel to the same output port. This design of WSS provides space and spectrum switching flexibility. Figure 3.1 illustrates two possible designs of an $N \times M$ WSS without contention. Figure 3.1(a) presents a design of an $N \times M$ WSS without contention as a single independent component while Figure 3.1(b) presents a design of an $N \times M$ WSS without contention as a combination of components which is implemented by coupling N inbound $1 \times M$ WSSs to M outbound $N \times 1$ WSSs in parallel.

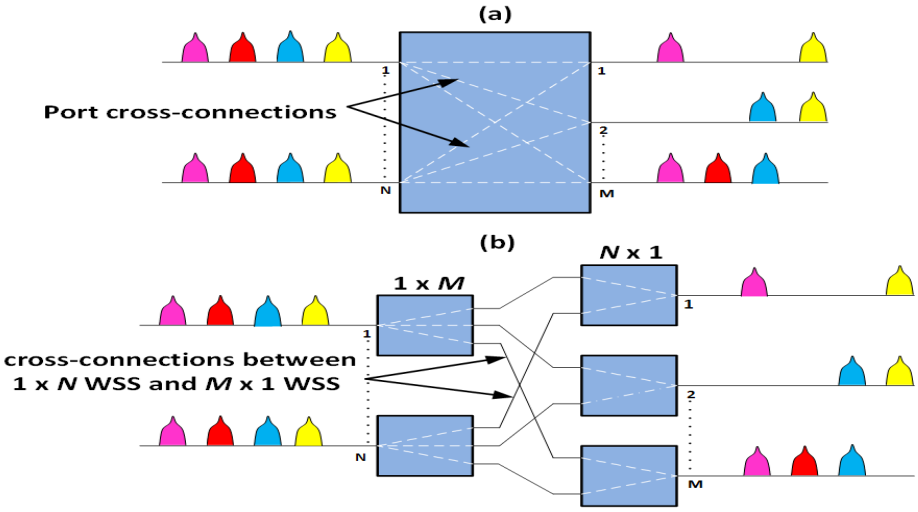

Figure 3.1: Designs of $N \times M$ WSSs without contention.

Assuming an $N \times M$ WSS without contention supports W number of wavelength channels where multiple wavelength channels can be switched or blocked between N input ports and M output ports. Suppose the number of copies of the same colour of wavelength channel that can be passed from N input ports to M output ports at the same time is a, the number of copies of the same colour of wavelength channel that can be blocked across N input ports is $(N - a)$, and the maximum number of copies of the same colour of wavelength channel that can be passed from N input ports to M output ports at the same time is x. The number of different ways that $(N - a)$ can be chosen from N input ports is $\binom{N}{N-a}$ and the number of different ways that a can be switched to M output ports without wavelength collisions is $\frac{M!}{(M-a)!}$. For all possible values of a and for W number wavelength channels, the total number of possible states of the $N \times M$ WSS without contention is

$$T = \left(\sum_{a=0}^{x} \frac{M!}{(M-a)!} \binom{N}{N-a} \right)^{W} \tag{3.4}$$

$$if\ N \leq M, x = N, elseif\ N > M, x = M$$

Thus, the flexibility of an $N \times M$ WSS without contention is

$$F(S) = Wlog \left(\sum_{a=0}^{x} \frac{M!}{(M-a)!} \binom{N}{N-a} \right) \tag{3.5}$$

$$if\ N \leq M, x = N, elseif\ N > M, x = M$$

Note that the flexibility measurement model in Equation.(3.5) can be utilized to measure the flexibility of any port dimension of WSS. Also, when $N \leq M$, $x = N$, this is because the maximum number of copies of the same of colour wavelength channel that can be passed from N input ports to M output ports at the same time without wavelength collisions at the output ports is N. Otherwise, when $N > M, x = M$, this is because the maximum number of copies of the same of colour wavelength channel that can be passed from N input ports to M output ports at the same time without wavelength collisions at the output ports is M. The derivation process is validated in Figure 3.2 with an illustration of the total number of possible states of a 2 × 3 WSS without contention which supports one colour of wavelength channel. When a = 0 (see Figure 3.2(a)), the WSS has only 1 state because wavelength-1 and wavelength-2 (which are the same colour of wavelength) are blocked at input port-1 and input port-2, respectively, i.e., no wavelength channel is successfully passed. When $a = 1$ (see Figure 3.2(b)), only one wavelength channel can be successfully passed at a time. Either wavelength-1 on input port-1 is switched to any of the 3 output ports while wavelength-2 on input port-2 remains blocked or wavelength-2 on input port-2 is switched to any of the 3 output ports while wavelength-1 on input port-1 remains blocked. Thus, the number of possible states when $a = 1$ is 6. Finally, when $a = 2$ (see Figure 3.2(c)), wavelength-1 and wavelength-2 can successfully be switched from input port-1 and input port-2, respectively, to any of the 3 output ports provided that wavelength collisions are avoided. The number of possible states when $a = 2$ is 6. Thus, the total number of possible states T is 13.

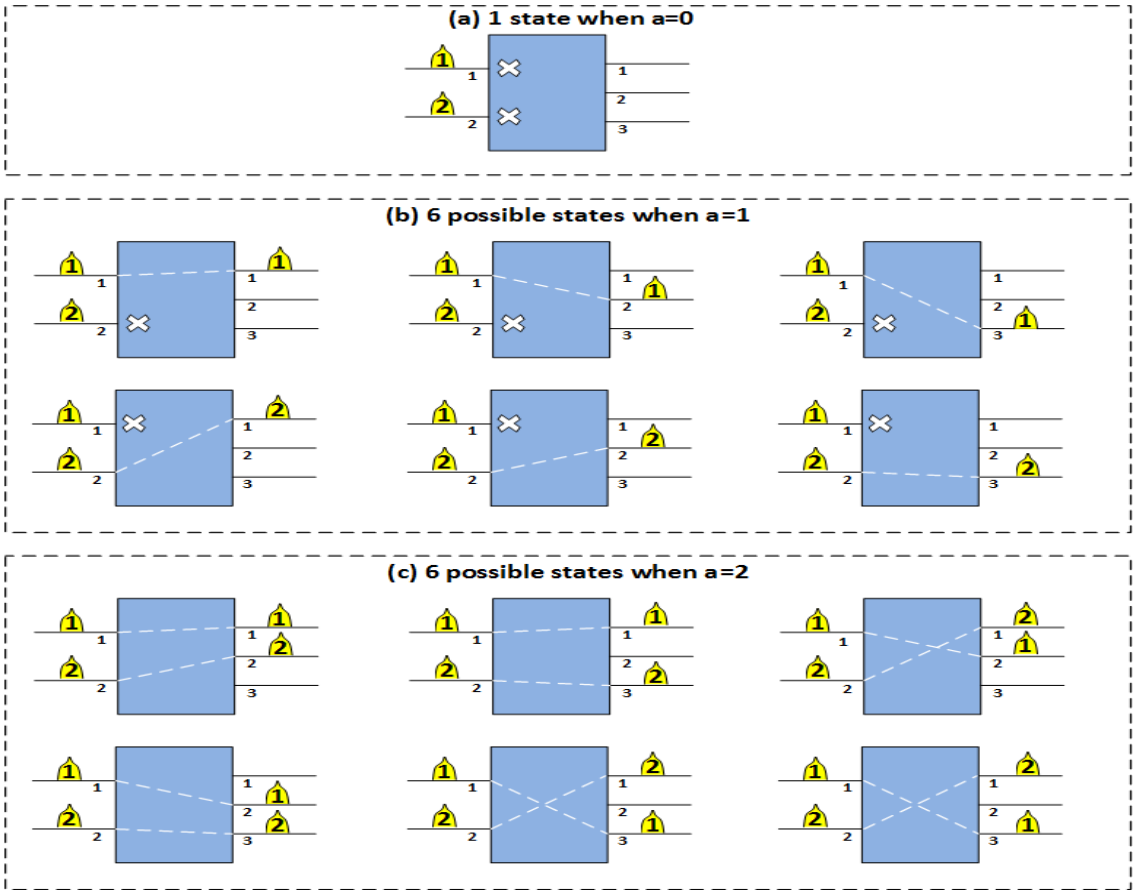

Figure 3.2: The total number of possible states of a 2 × 3 WSS without contention which supports one colour of wavelength channel.

Using Equation.(3.4), the total number of possible states of the 2 × 3 WSS is calculated as follows:

- Number of states when $a = 0$ is

$$\binom{N}{N-a} \times \frac{M!}{(M-a)!} = \binom{2}{2-0} \times \frac{3!}{(3-0)!} = 1$$

- Number of states when $a = 1$ is

$$\binom{N}{N-a} \times \frac{M!}{(M-a)!} = \binom{2}{2-1} \times \frac{3!}{(3-1)!} = 6$$

- Number of states when $a = 2$ is

$$\binom{N}{N-a} \times \frac{M!}{(M-a)!} = \binom{2}{2-2} \times \frac{3!}{(3-2)!} = 6$$

The total number of possible states $T = (1 + 6 + 6)^1 = 13$. Thus, the total number of states calculated from Equation.(3.4) is the equal to the total number of possible states displayed in Figure 3.2.

In order to evaluate the relationship between the flexibility of a WSS without contention and design features such as spectral range and port dimension, the measured flexibility of a 20 port WSS as a function of different port dimensions and number of wavelength channels is illustrated in Figure 3.3. Several insights are noted from Figure 3.3. Firstly, it can be noted that the flexibility of each port dimension configuration increases as the number of wavelength channels increases. This is due to the fact that more connection states can be establish as the spectral range of the WSS increases. Secondly, it is observed that from 9 to 100 wavelength channels, as the equilibrium between N input ports and M output ports reduces, the flexibility of the WSS also reduces as a result of lower port cross-connections between input ports and output ports. The 10×10 port dimension demonstrated the highest flexibility while 1×19 port dimension demonstrated the lowest flexibility. These observations show that port cross-connection and spectrum range are important design features for determining the resultant flexibility and connectivity of an $N \times M$ WSS without contention.

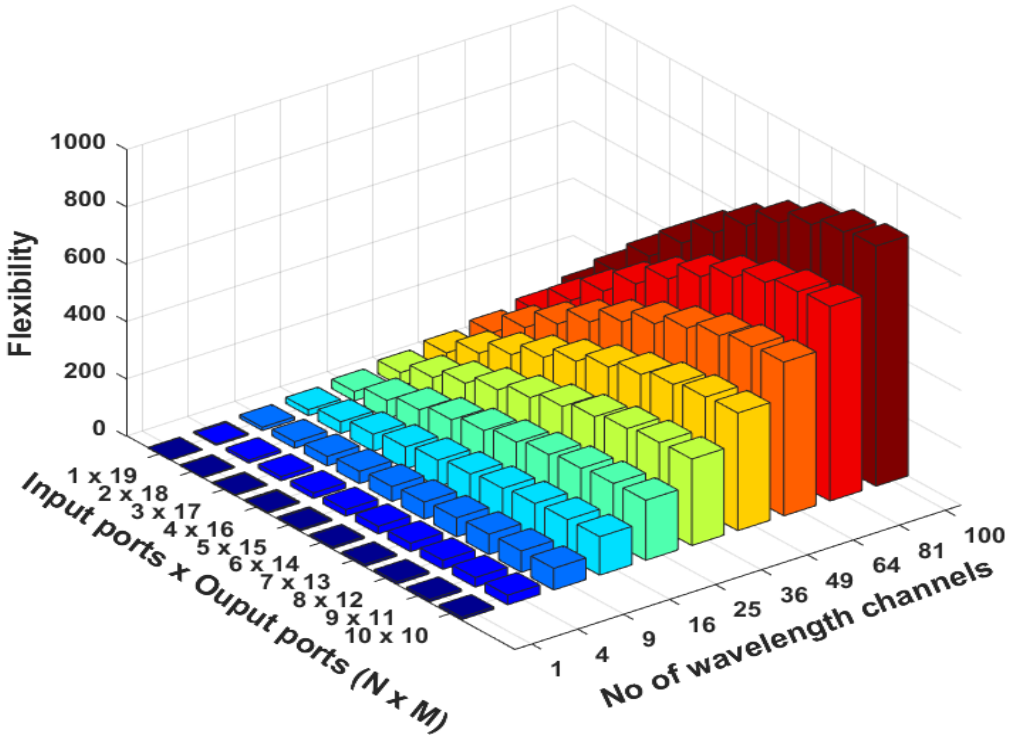

Figure 3.3: Flexibility of different port dimensions of a 20 port WSS.

3.3.2 $N \times M$ WSS with Contention

Figure 3.4 illustrates an $N \times M$ WSS design with contention which is implemented by connecting an inbound $N \times 1$ WSS to an outbound $1 \times M$ WSS in series. In this design, multiple copies of the same colour of wavelength channel cannot be simultaneously passed from multiple input ports to multiple output ports. This is because the two WSSs are connected through a single fibre. Thus, in comparison to the $N \times M$ WSS design without contention (Figure 3.1), the flexibility and connectivity of this design of WSS is drastically reduced.

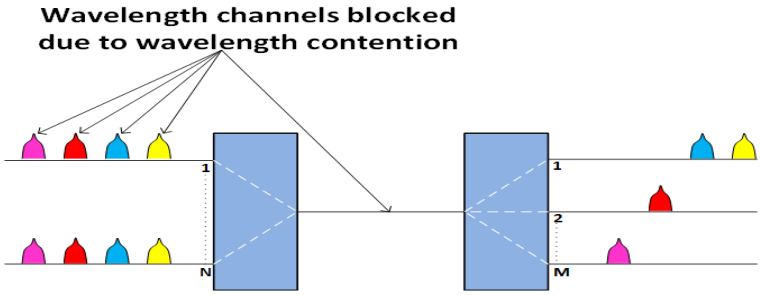

Figure 3.4: Designs of an $N \times M$ WSS with contention.

The flexibility measurement model for an $N \times M$ WSS with contention is derived in a similar way to the $N \times M$ WSS without contention. The difference is that the maximum number of copies of the same colour of wavelength channel than can be passed from N input ports to M output ports at the same time is equal to one (i.e., $x = 1$). Therefore, the flexibility measurement model of an $N \times M$ WSS with contention is

$$F(S) = Wlog\left(\sum_{a=0}^{1} \frac{M!}{(M-a)!}\binom{N}{N-a}\right) \tag{3.6}$$

Equation. (3.6) can be utilized to measure the flexibility of any port dimension of WSS. In order to evaluate the effect of wavelength contention on flexibility, Figure 3.5 presents a comparison of the flexibility of a 4 × 16 WSS without contention and a 4 × 16 WSS with contention as function of number of wavelength channels. As expected, the 4 × 16 WSS without contention demonstrates a higher flexibility than the 4 × 16 WSS with contention across the different number of wavelength channels. Particularly, it is noted that as the number of wavelength channels increases, the difference in flexibility between the 4 × 16 WSS without contention and the 4 × 16 WSS with contention increases. This is because the $N \times M$ WSS design without contention allows multiple copies of the same colour of wavelength channel to

be simultaneously passed from multiple input ports to multiple output ports whereas the $N \times M$ WSS design with contention does not allow. Thus, in $N \times M$ WSS design with contention, there is internal wavelength blocking and the use of wavelength resources are restricted. Therefore, the $N \times M$ WSS design without wavelength contention provides greater flexibility, connectivity and more efficient use of wavelength resources.

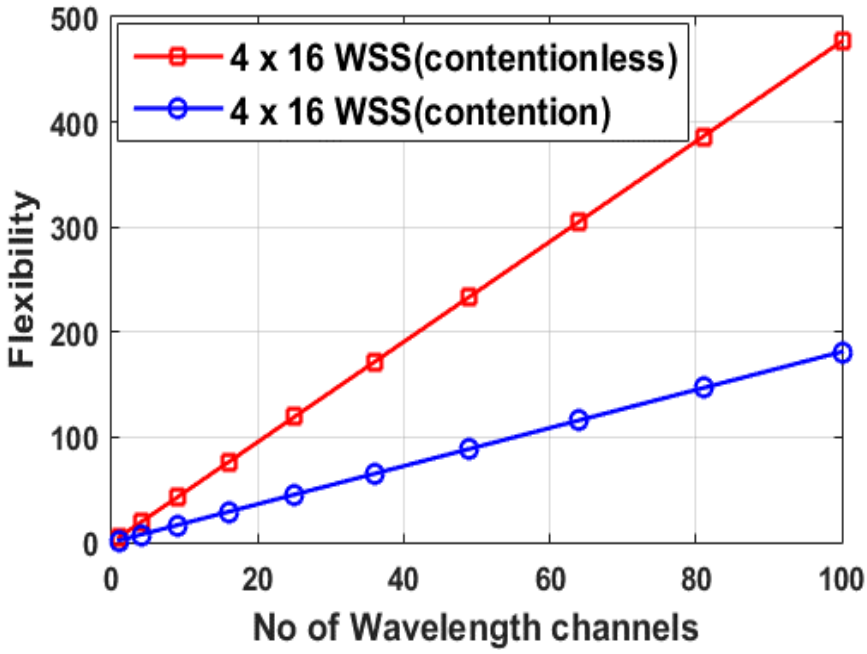

Figure 3.5: Comparison of the flexibility of a 4 × 16 WSS without/with contention.

3.3.3 $N \times M$ WSS with I/O Port Dimension Reconfigurability

A design of $N \times M$ WSS with I/O port dimension reconfigurability has the ability to rearrange its port dimensions to meet different switching/routing requirements. Thus, this WSS design provides architectural, spectrum switching and space switching flexibility. Figure 3.6 illustrates a 10 port WSS with port dimension reconfigurability rearranging its port dimension connection state from a 2 × 8 port dimension to connection state of a 4 × 6 port dimension.

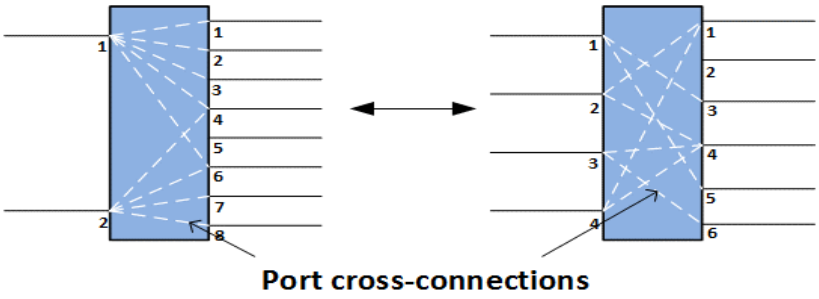

Figure 3.6: An $N \times M$ WSS with I/O port dimension reconfigurability.

To derive the flexibility measurement model, it is assumed that the total number of ports $P = N + M$. Therefore, $M = P - N$. Assuming an $N \times M$ WSS without contention which supports W number of wavelength channels, can switch or block a single colour or multiple colours of wavelength channels between N input ports and M output ports, and can independently be reconfigured to different configurations of input to output ports dimensions. The flexibility is

$$F(S) = log\left(\sum_{N=1}^{(P-1)}\left(\left(\sum_{a=0}^{x}\frac{(P-N)!}{((P-N)-a)!}\binom{N}{N-a}\right)^{W}\right)\right) \quad (3.7)$$

$$if\ N \leq P-N, x = N\ and\ if\ N > P-N, x = P-N$$

Note that the possible range of values of the index summation of N depends on the step size of the WSS. For example, a 10 port WSS with a I/O port dimension reconfigurability step size of 1 will have index summation values of {1, 2, 3, 4, 5, 6, 7, 8, 9} which represent the following port dimensions {1 × 9, 2 × 8, 3 × 7, 4 × 6, 5 × 5, 6 × 4, 7 × 3, 8 × 2, 9 × 1}. In another example, a 10 port WSS with a I/O port dimension reconfigurability step size of 2 will have index summation values {1, 3, 5, 7, 9} which represent the following port dimensions {1 × 9, 3 × 7, 5 × 5, 7 × 3, 9 × 1}. Additionally, the flexibility of a WSS design with wavelength contention and I/O port dimension reconfigurability is modelled in a similar way, the only difference is that the maximum number of copies of the same colour of wavelength channel that can be successfully passed from N input ports to M output ports at the same time is one (i.e., $x = 1$).

In order to evaluate the impact of port reconfigurability feature on the flexibility of a WSS, a comparison of the flexibility of a non-configurable 1 × 19 WSS, a non-configurable 10 × 10 WSS and four 20 ports WSSs with port dimension reconfigurability step sizes of 1, 2, 4 and 8, as function of the number of wavelength channels is shown in Figure 3.7. The following observations are noted from Figure 3.7. Firstly, at 25 wavelength channels, among the WSSs

with I/O port dimension reconfigurability, the WSS with step size 1 demonstrates the highest flexibility due to its ability to be reconfigured to the most number of possible port dimensions, while the non-configurable 1 × 19 WSS demonstrates the lowest flexibility because it has the lowest number of port cross-connections. Secondly, the non-configurable 10 × 10 WSS demonstrates a lower flexibility than the reconfigurable 20 port WSSs with step sizes 2, 4 and 8 due to its static port configuration, but offers a higher achievable port connectivity (port connectivity is defined as the number of port cross-connection between input and output ports and is equal to the product of the number of input ports and output ports). This is due to the fact that the maximum achievable port cross-connections from all possible port dimensions of the reconfigurable 20 port WSSs with step sizes 2, 4 and 8 are always lower than the 10 × 10 WSS. This analysis shows that the port reconfigurability step size is important in determining the resultant flexibility and connectivity of a WSS.

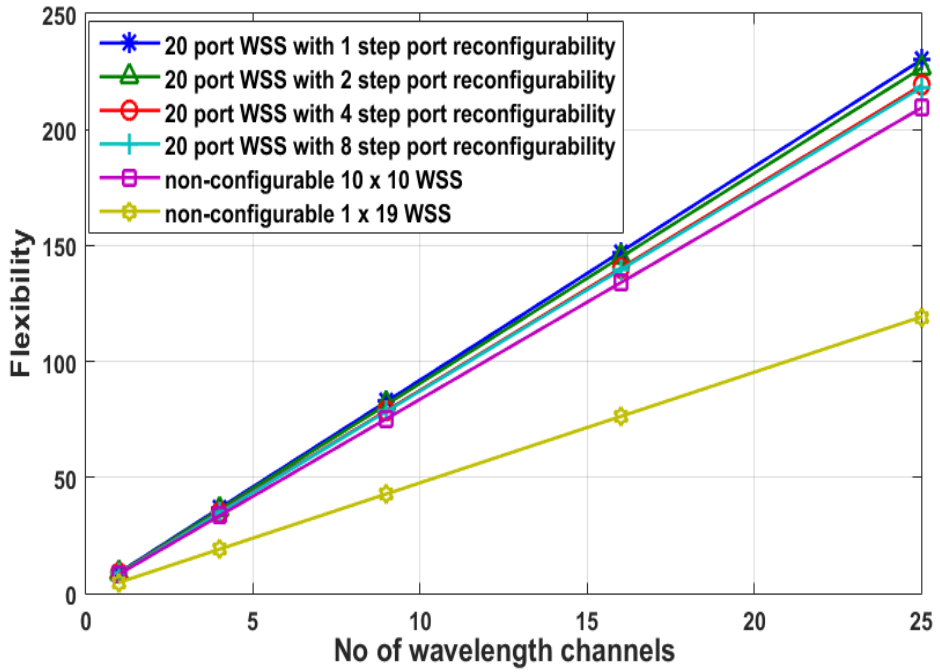

Figure 3.7: Comparison of the flexibility a non-configurable 1 × 19 WSS, a non-configurable 10 × 10 WSS, and four 20 port WSSs with different port reconfigurability step sizes of 1, 2, 4 and 8.

3.4 Flexibility Measurement of Spectrum Selective Switch

3.4.1 $N \times M$ SSS without Contention

An $N \times M$ SSS without contention can independently switch or block multiple spectral slots of different widths across multiple input ports and output ports. Compared to the WSS without contention, this device provides space switching and finer spectral switching flexibility. Figure 3.8 illustrates an $N \times M$ SSS without contention as a single component. The $N \times M$ SSS design without contention can also be implemented by coupling N inbound $1 \times M$ SSSs to M outbound $N \times 1$ SSSs in parallel. In flexible grid networks, an optical signal may occupy one or more spectral slots or infinite number of combinations of spectral slot sizes. Thus, because of the complexity and dynamic nature of flexible grid networks, a spectral granularity factor is used in this thesis. If the spectrum of a fixed grid network has a channel spacing of 50 GHz and the spectrum of a flexible grid network has a spectral slot spacing of 12.5 GHz. For every wavelength channel in the fixed grid network, it is assumed there are 4 spectral slots in the flexible grid network, i.e., the spectral granularity factor of the SSS is 4.

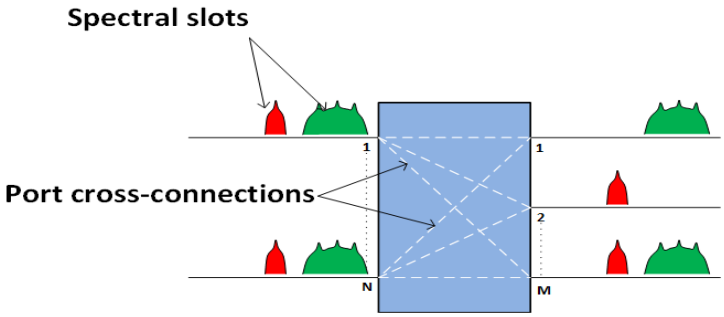

Figure 3.8: Design of an $N \times M$ SSS without contention.

In order to derive the flexibility measurement model of the $N \times M$ SSS without contention, the following assumption are made: a is the number of copies of the same colour and size of spectral slots that can be passed from N input ports to M output ports at the same time, $N - a$ is the number of copies of the same colour and size of spectral slots that are blocked, x is the maximum number of copies of the same colour and size of spectral slots that can be passed from N input ports to M output ports at the same time, k is the spectral granularity factor, and kW is the number of spectral slots that the SSS can support. Suppose an $N \times M$ SSS can independently switch or block kW spectral slots between N input ports and M output ports. The flexibility is

$$F(S) = kWlog\left(\sum_{a=0}^{x} \frac{M!}{(M-a)!}\binom{N}{N-a}\right) \tag{3.8}$$

$$if\ N \leq M, x = N\ and\ if\ N > M, x = M$$

Equation.(3.8) can be utilized to measure any port dimension of SSS. To evaluate the impact of spectrum switching on flexibility, and design trade-off between flexibility and port connectivity, Figure 3.9 depicts the measured flexibility of various port dimensions of WSS and SSS without contention as a function of different number of wavelength channels. A spectral granularity of 12.5GHz was used to calculate the flexibility of the SSS. It can be noted from Figure 3.9 that for each port dimension pair of the WSS and SSS, the flexibility increases as the number of wavelength channels increases. Furthermore, it is noted that the increase in flexibility has a greater effect on the SSS than the WSS, this is due to the finer spectral switching granularity of the SSS. Also, trade-offs between flexibility and port connectivity of the WSS and SSS can be observed. For example at 100 wavelength channels, the 10 × 11 WSS has a greater port connectivity than the 5 × 6 SSS but with a lower flexibility. The same trade-off exists between the 5 × 6 WSS and the 2 × 3 SSS. Therefore, such design trade-offs and analysis provide information on different levels of flexibility and port connectivity for the design of optical subsystems and ROADMs.

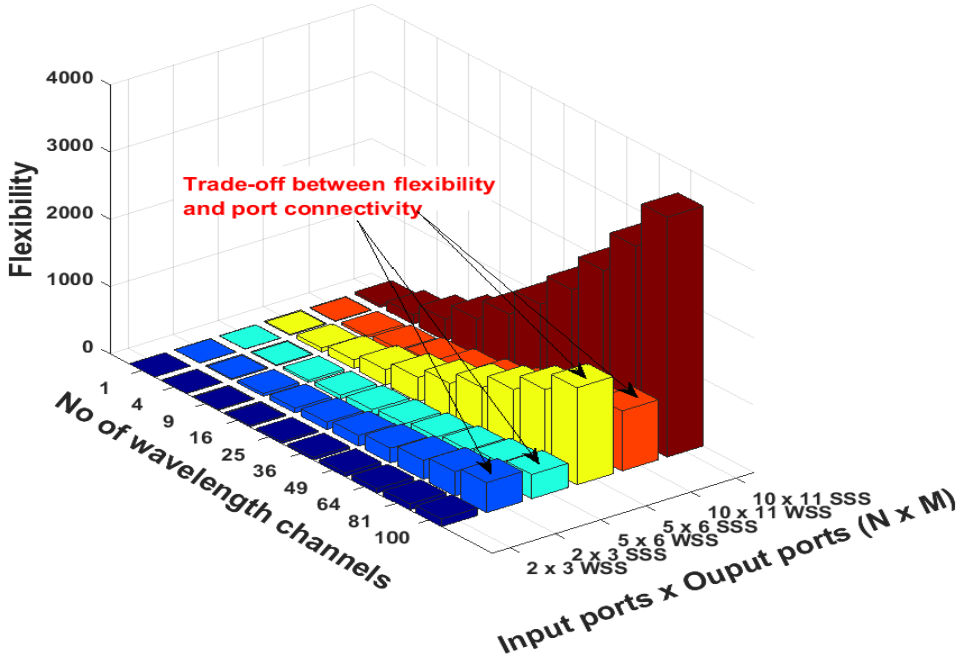

Figure 3.9: Comparison of the flexibility of different port dimensions of WSSs/SSSs.

The difference in flexibility between the WSS and SSS is further evaluated based on the maximum port dimension of a commercially available WSS/SSS, which is to the best of my knowledge a 1 × 35 WSS/SSS [39]. Figure 3.10 displays the measured flexibility of a WSS and SSS with a 1 × 35 port dimension as a function of wavelength channels. The flexibility of SSS was calculated with spectral granularity of 12.5GHz. It can be observed from Figure 3.10 that the SSS demonstrates higher flexibility than the WSS across all the different number of wavelength channels. This is due to the spectral granularity of the SSS. In particular, at a spectral range of 100 wavelength channels, the SSS demonstrates about 300% higher flexibility than the WSS.

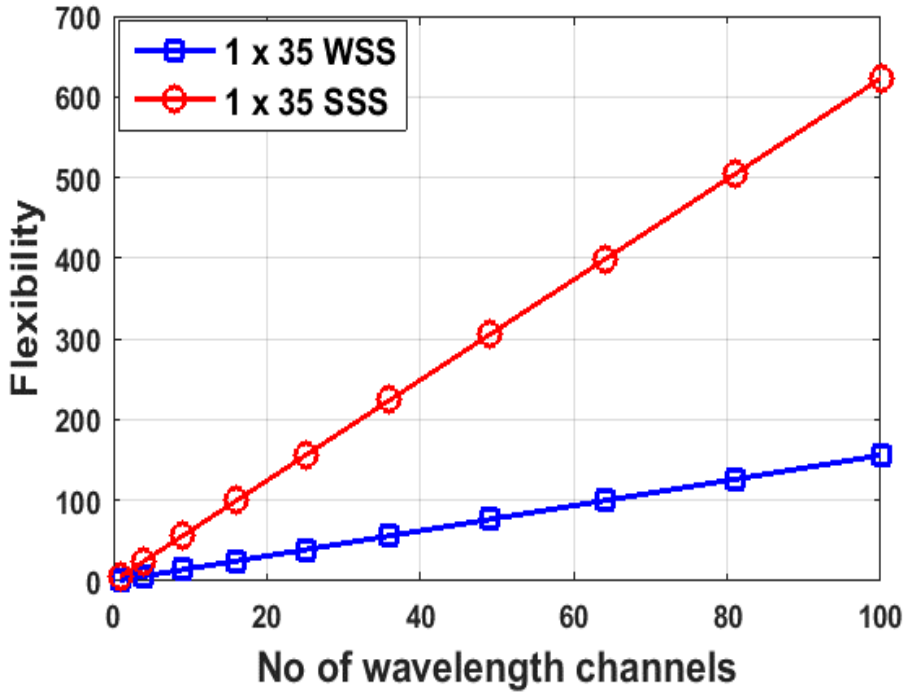

Figure 3.10: Comparison of the flexibility of a 1 × 35 WSS/SSS without contention.

3.4.2 $N \times M$ SSS with Contention

The flexibility measurement model of the $N \times M$ SSS design with contention is derived in a similar way to the $N \times M$ WSS design with contention, the only difference is the spectral granularity factor k of the SSS. Thus, the flexibility is

$$F(S) = kWlog\left(\sum_{a=0}^{1} \frac{M!}{(M-a)!}\binom{N}{N-a}\right) \tag{3.9}$$

3.4.3 $N \times M$ SSS with I/O port Dimension Reconfigurability

An $N \times M$ SSS with I/O port dimension reconfigurability provides port dimension reconfigurability in addition to finer spectrum switching granularity. This SSS design has a greater level of flexibility in comparison to the previously discussed $N \times M$ WSS and $N \times M$ SSS designs. The flexibility measurement model of the $N \times M$ SSS design with port reconfigurability is modelled in a similar way to the $N \times M$ WSS design with port reconfigurability but with a spectral granularity factor k. Thus, assuming this SSS design does not have contention, the flexibility is

$$F(S) = log\left(\sum_{N=1,}^{(P-1)}\left(\left(\sum_{a=0}^{x}\frac{(P-N)!}{((P-N)-a)!}\binom{N}{N-a}\right)^{kW}\right)\right) \quad (3.10)$$

$$if\ N \leq P-N, x = N\ and\ if\ N > P-N, x = P-N$$

3.5 Chapter Conclusion

In this chapter, a flexibility measurement methodology for optical networks was reported. Using this methodology, flexibility measurement models for different design configurations of WSS and SSS were derived and proposed. Also, an example to validate the derivation process was presented. Furthermore, the flexibility of different configurations of WSS and SSS were measured, and design trade-offs between flexibility, connectivity and design features were evaluated.

Numerous significant insights for the design of WSS and SSS were noted from the results and analysis reported in this chapter. First, it was observed that design features such as: contention, spectral range, switching dimensions (space and spectrum), port dimension, I/O port dimension reconfigurability and spectral granularity determine the resultant performance of WSS and SSS. Also, the selection and combination of these design features offer different levels of flexibility and performance. Hence, an WSS/SSS can be designed to deliver a set of functions or a certain performance. Figure 3.3 showed that the flexibility of a WSS increases with higher port cross-connections and spectral range, and Figure 3.5 showed that a WSS without contention has a greater amount of flexibility than a WSS with contention due to more efficient use of wavelength resources. Hence, if an $N \times M$ WSS design is required to build an optical node with high connectivity, flexibility and node degree scalability; the port dimension size, spectral range and wavelength contention are important design features to be considered. A

high level of port cross-connection and an equilibrium distribution between the number of input ports and output ports increases the connectivity and switching flexibility of a WSS, this is because multiple network paths between input ports (source) and output ports (destination) can established. Also, for node degree scalability, the port dimension of the device is important, this is because there must be enough input/output ports provisions for node degree expansion, i.e., adding more node degrees to an existing mesh network or expanding the add/drop network of an optical node. Furthermore, a WSS without contention improves connectivity of add/drop networks of optical nodes as the selection of wavelength resources are not restricted and internal wavelength blocking is eliminated.

Another important design feature for a WSS/SSS design is port dimension reconfigurability and spectral granularity. Figure 3.7 showed that a WSS with different steps of port dimension reconfigurability offers greater flexibility than a non-configurable WSS. If a WSS design is required to manage varying and unpredictable bidirectional traffic flows, port dimension reconfigurability is an important design feature because the WSS can adjust is structure to respond to the varying traffic. Spectral granularity is another important design feature for a WSS/SSS. Figure 3.9 and Figure 3.10 show that an SSS has more flexibility than WSS, this is due to spectral granularity of the SSS. Therefore, when designing flexible grid networks, spectral granularity is an important design feature because the switching components in the network should be able to provide finer spectrum switching flexibility for optical signals of variable sizes.

4 Bandwidth Variable Transponders and Optical Subsystems

4.1 Introduction

In this chapter, the flexibility measurement models of different configurations of optical transmitters ranging from single optical carrier to multiple optical carriers with fixed/programmable transmission features are derived and proposed. Furthermore, the design trade-offs between the flexibility, other KPIs and design features of different BVT configurations are evaluated. Next, the flexibility measurement models of different optical switching and transmission subsystems are derived and proposed, and the design trade-offs between the flexibility, other KPIs and design features of different optical switching and transmission subsystems are also evaluated. Numerous significant insights from the results and theoretical analysis reported in this chapter are highlighted. All flexibility measurements models used to measure the flexibility of BVTs and subsystems in this chapter were implemented in MATLAB. The different BVT modules in Table 4.3 and cost model in Equation.(4.16) presented in this chapter was contributed by Matthias Gunkel of Deutsche Telekom.

4.2 Flexibility Measurement Models for Bandwidth Variable Transponders

This section examines a variety of transponder configurations which provide different degrees of flexibility and performance. Table 4.1 displays an extensive range of optical transponder configurations equipped with different functionalities, and Table 4.2 presents the list and definition of all parameters used in the derivation of all flexibility measurement models for optical transponders.

Table 4.1: Summary of different transponder configurations

Transponder Configuration	Laser Type	Optical Carrier Category	Programmable Features
1	Fixed grid non -tunable	Single carrier	All parameters are fixed
2	Fixed grid tunable laser	Single carrier	Modulation formats, Symbol rates, and Wavelength channels

3	Flexible grid tunable laser	Single carrier	Modulation formats, Symbol rates, and Spectral slots
4	Fixed grid tunable laser	Single carrier	Modulation formats, Symbol rates, Wavelength channels, and Electrical subcarriers
5	Flexible grid tunable laser	Single carrier	Modulation formats, Symbol rates, Spectral slots, and Electrical subcarriers
6	Fixed grid tunable laser	Multi carrier	Modulation formats, Symbol rates, Wavelength channels, and Optical carriers
7	Flexible grid tunable laser	Multi carrier	Modulation formats, Symbol rates, Spectral slots, and Optical carriers
8	Fixed grid tunable laser	Multi carrier	Modulation formats, Symbol rates, Wavelength channels, Optical carriers, and Electrical subcarriers
9	Flexible grid tunable laser	Multi carrier	Modulation formats, Symbol rates, Spectral slots, Optical carriers, and Electrical subcarriers

Table 4.2: Parameter definition for flexibility measurement models of optical transponders

Parameter	Definition
N	Number of BVTs in a subsystem
W	Number of tunable wavelength channels
k	Spectral granularity factor
kW	Number of tunable spectral slots
D	Total number of optical carriers
u	Number of optical carriers transmitting at the same time
$D-u$	Number of optical carriers that are off
B	Number of programmable modulations formats
E	Number of programmable symbol rates
r	Number of programmable electrical subcarriers

Configuration 1: This configuration is the least flexible transmitter design, which is a single optical carrier transmitter with fixed capacity, i.e., all parameters are fixed. The transmitter consists of a single non-tunable laser, supports one modulation format, and one symbol rate on fixed grid networks. This transmitter design provides no channel flexibility as it supports only one bitrate. As illustrated in Figure 4.1, this transmitter configuration has only two states, i.e.,

when the transmitter is transmitting an optical channel and when the transmitter is off (no transmission).

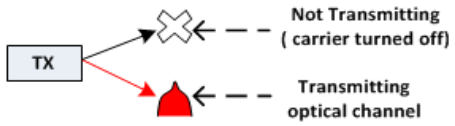

Figure 4.1: Possible states of configuration 1.

The total number of possible states of this transmitter configuration $T = 2$ and the flexibility is

$$F(S) = \log(2) \tag{4.1}$$

Configuration 2: This is a single optical carrier transmitter which supports multiple bitrates with tunable wavelength channels, i.e., the transmitter can independently be configured to different modulation formats, symbol rates and tunable wavelength channels on fixed grid networks. This transmitter configuration provides channel flexibility and can adapt to different transmission requirements. Assuming that an optical channel can be independently programmed to B modulation formats, E symbol rates, and W tunable wavelength channels, the total number of possible states is

$$T = BEW + 1 \tag{4.2}$$

Therefore, the flexibility is

$$F(S) = \log(BEW + 1) \tag{4.3}$$

Figure 4.2 illustrates the possible states of a fixed grid transmitter with programmable modulations formats, programmable symbol rates, and tunable wavelength channels.

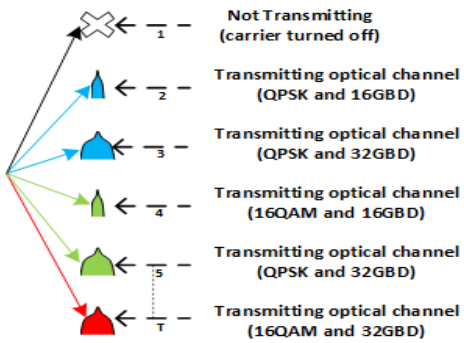

Figure 4.2: Possible states of configuration 2.

It should be noted that the model in Equation.(4.3) can be used to measure the flexibility when not all parameters are programmable. For example, a transmitter that supports multiple bitrates with programmable modulation formats but at a fixed symbol rate and wavelength channel. In this transmitter design, different modulation formats are achievable at a fixed symbol rate and wavelength channel, therefore the value of E and W is one and the total number of possible states is

$$T = B + 1 \tag{4.4}$$

Configuration 3: This configuration is a single carrier optical transmitter which supports multiple bitrates on flexible grid networks. This transmitter configuration has similar design features to the configuration 2 transmitter. However, the difference is that the fixed grid tunable laser is upgraded to a flexible grid tunable laser with finer tuning and spectral granularity. This configuration is modelled in a similar way to configuration 2 but with a spectral granularity factor of the laser. Suppose, the transmitter can be configured to kW spectral slots (over the same spectral range with configuration 2), and programmed to B modulation formats and E symbol rates. The flexibility is

$$F(S) = \log(BEkW + 1) \tag{4.5}$$

Configuration 4: This configuration is a single optical carrier transmitter where multiple orthogonal electrical subcarriers with different combinations of modulation formats and symbol rates can be created in the electrical domain and injected onto an optical carrier on fixed grid networks. In addition to channel flexibility, finer bitrate granularity may be realized. Suppose r electrical subcarriers can be independently configured to B modulation formats, E symbol rates, and injected onto any W wavelength channels, then the flexibility is

$$F(S) = log\left(\left(\sum_{r=1}^{r}(BE)^r W\right) + 1\right) \tag{4.6}$$

Configuration 5: This configuration is a single optical carrier transmitter design similar to configuration 4 but in flexible grid networks. This transmitter design provides spectral granularity in addition to finer bitrate granularity. Therefore, multiple subcarriers with different modulation formats and symbol rates are electrically generated and injected on kW tunable spectral slots. The flexibility of this configuration is

$$F(S) = log\left(\left(\sum_{r=1}^{r}(BE)^r\, kW\right) + 1\right) \tag{4.7}$$

Configuration 6: This is a multicarrier optical transmitter design which supports multiple bitrates on fixed grid networks. Figure 4.3 presents a schematic of a sliceable, tunable, bandwidth variable yet fixed grid transmitter.

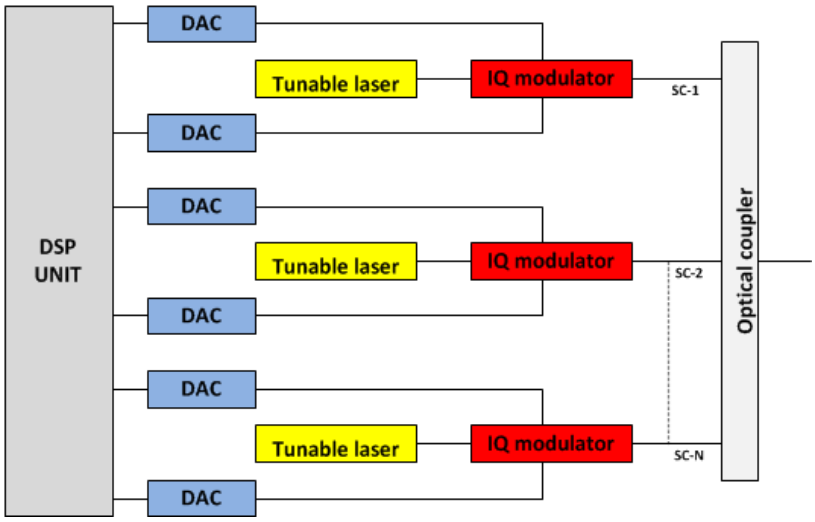

Figure 4.3: Multicarrier optical transmitter.

This transmitter configuration consists of tunable lasers, modulators, DACs and a DSP unit. Different number of optical carriers can be generated and combined to form super channels or multiple flows of optical paths of different bandwidths by turning on and turning off carriers, which in turn can be transmitted to a single destination or sliced to multiple destination, i.e., the transponder is sliceable. Figure 4.4 illustrates some possible states of a 3-optical carrier transmitter with programmable features. State-1 shows optical carrier one is active while optical carrier 2 and 3 are off. State-2 shows optical carrier 1 and 3 transmitting to form a super channel while optical carrier 2 is off. State-T shows optical carrier 1, 2 and 3 transmitting optical channels of different bandwidths to form multiple optical flows which are to be sliced to different destinations.

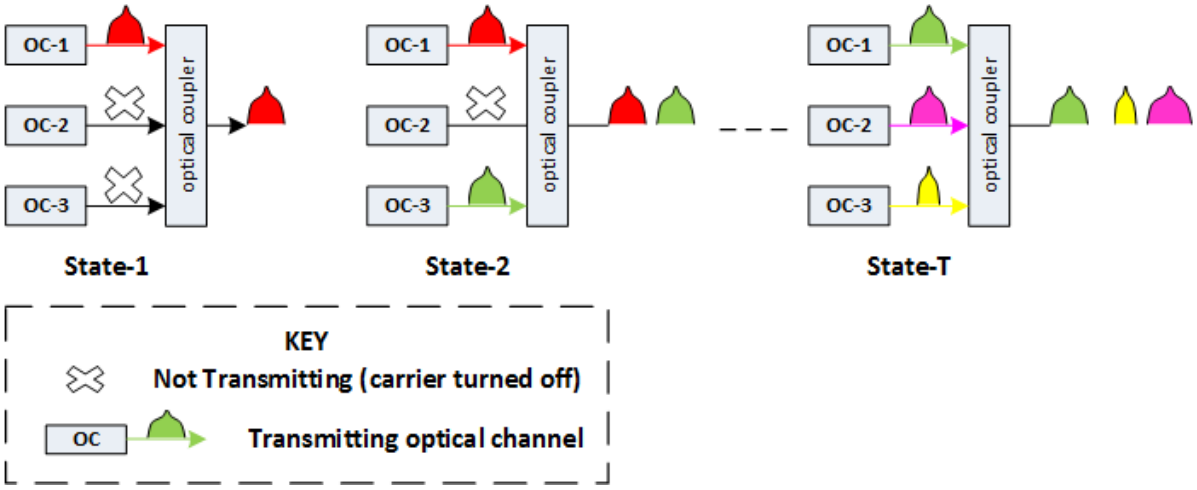

Figure 4.4: Possible states of a multicarrier optical transmitter.

Suppose D is the number of optical carriers of the transmitter, u is the number of optical carriers that are transmitted at the same time, and $(D-u)$ is the number of optical carriers that are off. The different number of possible ways that $(D-u)$ can be selected from D optical carriers is $\binom{D}{D-u}$. Suppose the different number of possible ways that u can be independently reconfigured to B modulations, E symbol rates and W tunable wavelength channels is $(BEW)^u$ (note that each carrier has the same number of programmable parameters), then the total number of possible states of the transmitter for all values of u is

$$T = \sum_{u=o}^{D} \left((BEW)^u \binom{D}{D-u} \right) \tag{4.8}$$

Thus, the flexibility of this transmitter configuration is

$$F(S) = log\left(\sum_{u=0}^{D} \left((BEW)^u \binom{D}{D-u} \right) \right) \tag{4.9}$$

It should be noted that the flexibility model for this transmitter design was derived without considering wavelength contention between optical carriers.

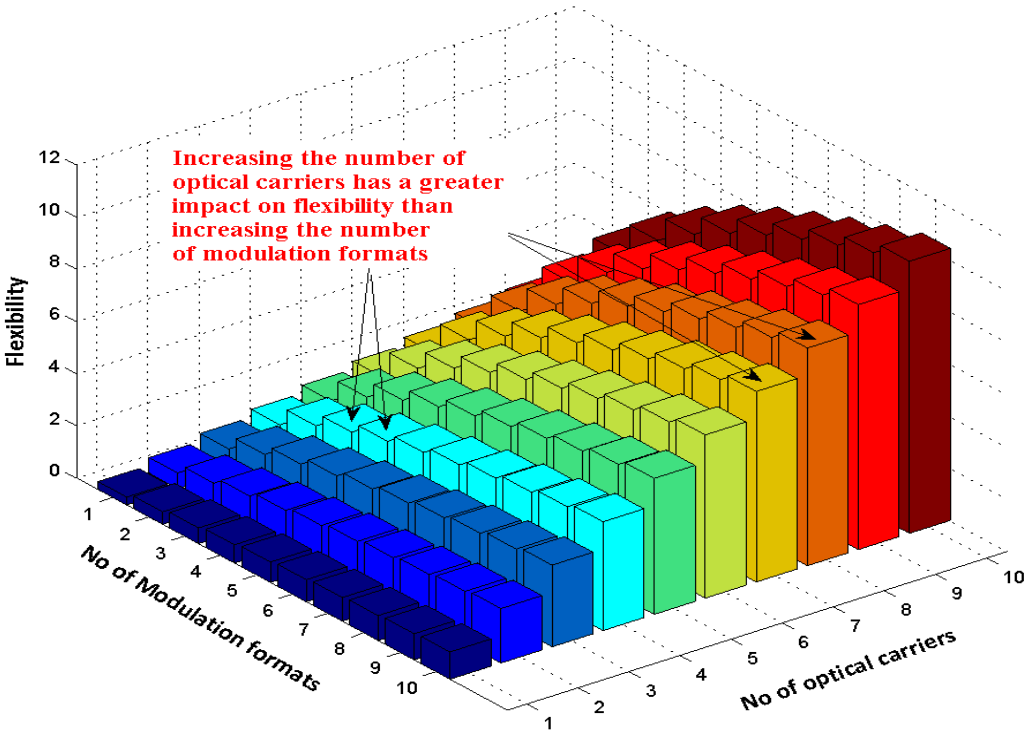

Figure 4.5: Impact of optical carriers and modulation formats on the flexibility of a transmitter.

To evaluate the impact of transmission features, i.e., optical carriers and modulation formats, on the flexibility of a transmitter, Figure 4.5 depicts the measured flexibility of a configuration 6 optical transmitter across different number of programmable modulation formats and optical carriers while other parameters are fixed (all other parameters are equal to one). In can be noted that both an increase in the number of programmable modulation formats and the number of optical carriers increases the flexibility of the transmitter. However, increasing the number of optical carriers has a greater impact on flexibility than increasing the number of programmable modulation formats. This is due to the fact that optical carriers provide more flexibility and connectivity in transmitter output. Such design features trade-offs are vital for optical transmitter design based on transmission requirements.

Configuration 7: This transmitter configuration is a multicarrier transmitter design equipped with flexible grid tunable lasers for flexible grid networks. Modelled in a similar way to configuration 6, assuming that each carrier can be independently tuned to kW spectral slots, B modulations and E symbol rates. The flexibility is

$$F(S) = log\left(\sum_{u=0}^{D}\left((BEkW)^{u}\binom{D}{D-u}\right)\right) \tag{4.10}$$

Note that the flexibility measurement model for this transmitter design was derived without considering spectral slot contention.

Configuration 8: This transmitter configuration is a multicarrier optical transmitter design where electrical subcarriers with different combinations of modulation formats and symbol rates can be generated and inserted onto different optical carriers to form super channels or multiple optical flows. Modelled in a similar way to configuration 6 but with an addition of r electrical subcarriers. The flexibility of this configuration of transmitter for fixed grid networks is

$$F(S) = log\left(\sum_{u=0}^{D}\left(\left(\sum_{r=1}^{r}(BE)^r W\right)^u \binom{D}{D-u}\right)\right) \tag{4.11}$$

The flexibility measurement model for all the previously presented fixed grid multicarrier optical transmitters were derived without considering wavelength contention between carriers. A flexibility measurement model to consider such contention is described below. To derive this model, the different number of possible ways that different W tunable wavelength channels can be selected without repetition for u optical carriers is calculated by $\frac{\boldsymbol{W}!}{(\boldsymbol{W}-\boldsymbol{u})!}$. Thus, the flexibility of configuration 8 transmitter without wavelength contention is

$$F(S) = log\left(\sum_{u=0}^{D}\left(\left(\sum_{r=1}^{r}(BE)^r\right)^u \frac{W!}{(W-u)!}\binom{D}{D-u}\right)\right) \tag{4.12}$$

It should be stated that Equation.(4.12) can only be applied when the number of tunable wavelength channels W is equal or greater than the number of optical carriers D ($W \geq D$), and that Equation.(4.11) and Equation.(4.12) are the generic flexibility measurement models for all transmitters configurations on fixed grid networks.

Configuration 9: This transmitter is similar to configuration 8, however the tunable lasers are upgraded to flexible grid tunable lasers with finer spectral tuning granularity. The flexibility is

$$F(S) = log\left(\sum_{u=0}^{D}\left(\left(\sum_{r=1}^{r}(BE)^r kW\right)^u \binom{D}{D-u}\right)\right) \tag{4.13}$$

The flexibility measurement model for all the previously presented flexible grid multicarrier

optical transmitters were derived without considering spectral slot contention between carriers. Modelled in a similar way to Equation.(4.12), the flexibility of configuration 9 transmitter without spectral slot contention is

$$F(S) = log\left(\sum_{u=0}^{D}\left(\left(\sum_{r=1}^{r}(BE)^{r}\right)^{u}\frac{kW!}{(kW-u)!}\binom{D}{D-u}\right)\right) \tag{4.14}$$

It should be stated that this model can only be applied when kW tunable spectral slot is equal or greater than the number of optical carriers D $(kW \geq D)$, and that Equation.(4.13) and Equation. (4.14) are the generic flexibility measurement models for all transmitters configurations on flexible grid networks.

4.3 Design Trade-offs of Bandwidth Variable Transponders

In order to evaluate the design trade-offs between different transmission features and KPIs, a comparative evaluation of the design features and KPIs of different BVTs modules presented in Table 4.3 which deliver rates between 100 Gb/s and 1 Tb/s is presented.

Table 4.3 Different BVT Modules

Module name	Normalized cost	No of carriers	Baud rate per carrier	Occupied bandwidth (GHz)	Reach (km)	Line Rate (Gb/s)	Modulation formats
CG-400-1L	9	1	32	37.5	38 116 536 2714	400 300 200 100	DP-256QAM DP-64QAM DP-16QAM DP-QPSK
CG-400-2L	11.7	2	32	75	536 2714 5429	400 200 100	DP-16QAM DP-QPSK DP-BPSK
CG-1T-5L	16.2	5	32	175	536 2714 5429	1000 500 250	DP-16QAM DP-QPSK DP-BPSK
CG-1T-4L	18.75	4	40	175	429 2171 4343	1000 500 250	DP-16QAM DP-QPSK DP-BPSK
CG-1T-3L	15	3	36	125	104 482 2443 4886	1000 666.7 333.3 166.7	DP-64QAM DP-16QAM DP-QPSK DP-BPSK

Figure 4.6(a)-(e) shows a visualization of the possible states and design features of each of BVT modules listed in Table 4.3. Each plot illustrates the trade-offs in line rates and lightpath

reach as a result of different adaptive transmitter features. Figure 4.6(f) shows a comparison of different KPIs (i.e., flexibility, maximum capacity, normalized cost, maximum spectral efficiency, connectivity, lightpath reach and through loss) for the various BVT modules presented in Table 4.3.

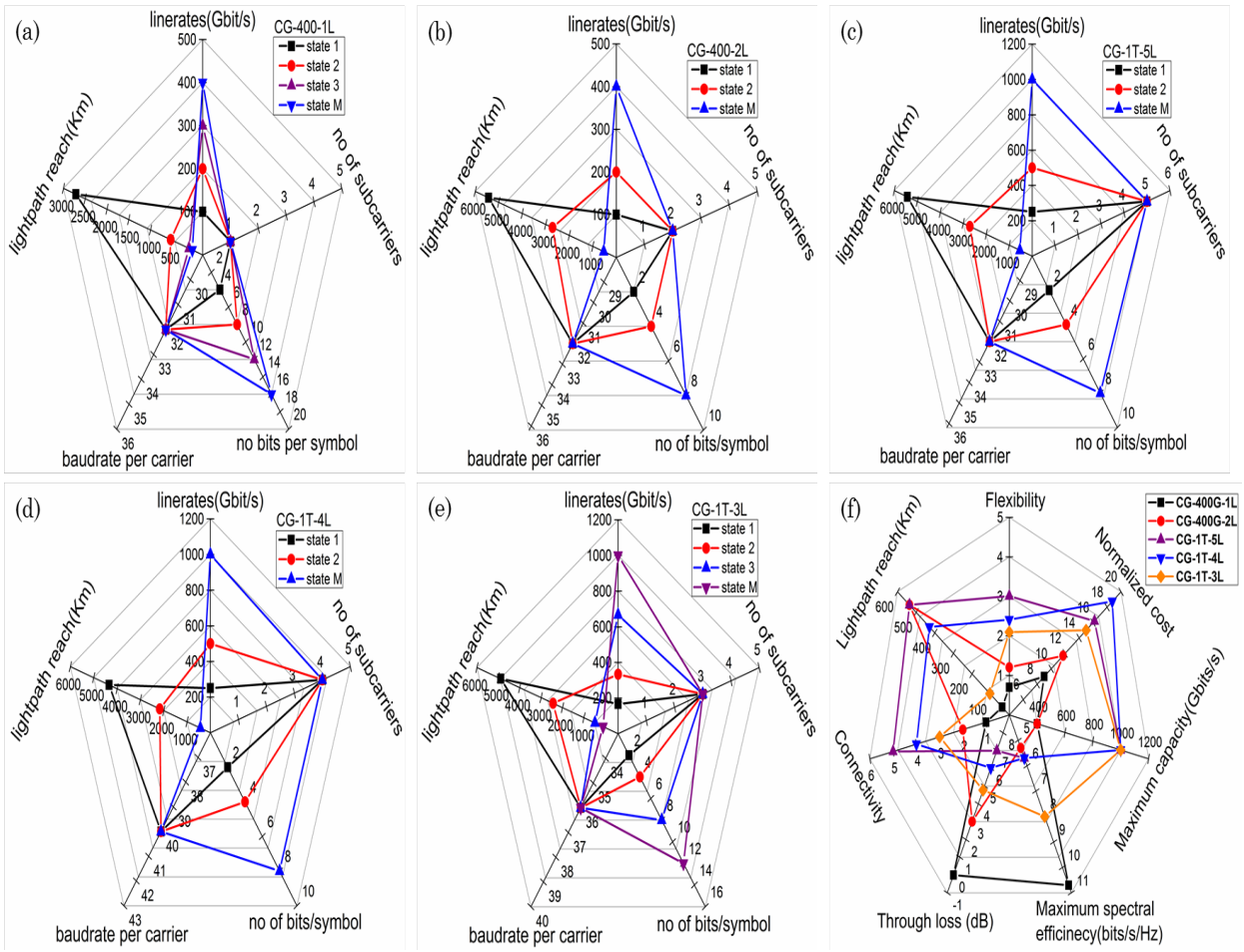

Figure 4.6: KPIs and design parameters of the various transponder molules in Table 4.3. (a) CG-400-1L. (b) CG-400-2L. (c) CG-1T-5L. (d) CG-1T-4L. (e) CG-1T-3L. (f) KPIs of the various trasnponder modules.

The connectivity of a BVT module is equal to the maximum number of optical carriers the BVT module can support. The spectral efficiency is calculated with a channel spacing of 50GHz. The through loss is calculated with Equation.(4.15), where L is equal to the total number of optical lasers connected to the optical coupler.

$$Through\ loss = -10log_{10}\left(\frac{1}{L}\right) \tag{4.15}$$

The normalized cost is scaled with respect to the number of lasers and baud rate as illustrated

in Table 4.4. The cost of the CG-400-1L BVT module is used as reference with a normalized cost value of 9. Y is equal to the percentage rise in cost of the number of lasers and R is equal to the percentage rise in the cost of the baud rate. Hence, Equation.(4.16) is used for calculating the normalized cost.

Table 4.4: Cost Scaling for BVT modules

No of lasers	Reference (%)	Baud rate	Reference (%)
1	100	32.0	100
2	130	35.6	111
3	150	40	125
4	167	42.7	133
5	180		

$$Normalized\ cost = \big((Y - 100\%) \times 9 + (R - 100\%) \times 9\big) + 9 \qquad (4.16)$$

CG-400G-1L: The CG-400G-1L module is the least expensive transponder module but has the lowest connectivity and flexibility when compared to the other BVT modules. This is because the module consists of only one optical laser and therefore can transmit only one optical channel to a single destination at a particular time. Despite the disadvantages introduced by a single optical carrier, the BVT module has no through loss because no optical coupler is needed for transmission operations. Furthermore, additional design trade-offs in maximum capacity, spectral efficiency and lightpath reach are noted when compared to the CG-400G-2L module. The CG-400-1L module provides the same maximum capacity as the CG-400L-2L module but with a greater spectral efficiency and a lower lightpath reach. This is due the fact that the CG-400-1L BVT has the highest spectral efficient programmable modulation format, i.e., 256QAM, among all the BVT modules.

CG-400L-2L: The CG-400L-2L is the second least expensive module and it has the second lowest flexibility and connectivity. Its low cost, flexibility and connectivity is attributed to the fact that it has only two lasers. This module has the same equivalent line rates with the CG-400-1L module but offers greater flexibility and connectivity because it has more programmable optical carriers. However, this comes at a demerit of a higher through loss due to the presence of a coupler. As highlighted in Figure 4.5, the number of optical carriers has a greater impact on flexibility than the number of programmable modulation formats. Such a trade-off is demonstrated between the CG-400L-1L module and the CG-400-2L module. CG-400G-2L has a higher number of optical carriers than CG-400G-1L but has a lower number of programmable modulation format. Therefore, CG-400G-2L demonstrates greater flexibility

than CG-400L-1L. Furthermore, in terms of spectral efficiency, the CG-400G-2L has the lowest maximum spectral efficiency among all BVT modules.

CG-1T-5L: The CG-1T-5L module has the highest flexibility and connectivity which is attributed to the fact that the BVT has the highest number of optical lasers. Therefore, it is the most flexible transponder module in terms of transmitter output and has the highest number of multi-flow optical channel combination. However, the merit in performance introduces a demerit of this module demonstrating the highest through loss. The CG-1T-5L module has the same line rates and maximum spectral efficiency as the CG-1T-4L module.

CG-1T-4L: The CG-1T-4L module is the most expensive module and this is due to the fact that it consists of 4 lasers and has a higher baud rate compared to all the other modules. From this module, it can be observed that the number of optical carriers as well as the DSP and electronics are important factors that determine the cost of a BVT module. This module has a higher flexibility and connectivity than the CT-1T-3L but with an equivalent maximum capacity.

CG-1T-3L: The CG-1T-3L module ranks in the middle in terms of flexibility, cost and connectivity. The CG-1T-3L has the same maximum capacity with CG-1T-5L and CG-1T-4L but with a greater maximum spectral efficiency. However, the merit in spectral efficiency introduces a drawback of a lower lightpath reach due to the BVT design having a higher spectral efficient modulation format which is DP-64QAM.

Significant insights into design trade-offs between different KPIs and design features for BVTs can be gained based on the KPIs and design features of the different transponder modules presented in Figure 4.6, and the theoretical analysis and results presented in this chapter. It can be observed from Figure 4.6(f) that connectivity scales directly with flexibility, i.e., the connectivity of a transponder module increases as the number of optical lasers increases and in turn the flexibility of the transponder module also increases. This proportional relationship also exists between connectivity and through-loss. This is because the higher the number of lasers present in the BVT module, the higher the port count of the optical coupler and in turn the higher the through loss. Furthermore, it was observed from Figure 4.6(f) that connectivity and flexibility relates directly with cost for the CG-1T-1L, CG-1T-2L and CG-1T-3L only. However, for the CG-1T-5L and CG-1T-4L, the CG-1T-5L offers a higher flexibility and connectivity than CG-1T-4L but at a lower cost. This is because the CG-1T-5L has a lower

baud rate electronic design. Capacity, spectral efficiency and lightpath reach do not demonstrate any direct relationship with any KPI. However, design trade-offs are noted when a certain performance or a combination of KPIs are required. For instance, if an operator requires a BVT module with a high line rate of 1 Tb/s. From Figure 4.6, it is observed that three BVT modules CG-1T-3L, CG-1T-4L and CG-1T-5L offer this performance requirement. It now depends on the trade-off or other KPIs which are important to the operator. For instance, if 1Tb/s is required with a high level of flexibility and connectivity while other KPIs are negligible, the CG-1T-5L fits this profile. If cost, through loss and spectral efficiency are important factors to deliver 1Tb/s, the CG-1T-3L fits this design requirement because it has the lowest cost, lowest through loss and the highest spectral efficiency at a demerit of lower lightpath reach. Alternatively, if an operator requires a BVT design with high spectral efficiency while other indicators are negligible, the CG-400G-1L fits the profile of this performance requirement. This is because it has the most spectral efficient modulation format. Also, the analysis described above helps understand the relationship between KPIs and design features of BVTs. The different transmitter configurations with various design features displayed Table 4.1 can aid network vendors/designers to select/design transmitters based on different degrees of flexibility and associated performance. For instance, if a single carrier transmitter with the flexible rate in a fixed grid network is required. The transmitter configuration will require a single fixed grid non-tunable laser with programmable modulation formats or programmable symbol rates or a combination of both depending on the performance required (e.g. variable spectral occupancy and spectral efficiency).

4.4 Flexibility Measurement Models for Optical Transmission and Switching Subsystems

This section presents the flexibility measurement models for optical transmission and switching subsystems. Table 4.5 presents the list and definition of all parameters used in the derivation of all flexibility models for optical transmission and switching subsystems presented in this chapter, and Figure 4.7 presents different optical subsystems discussed in this chapter.

Table 4.5: Parameter definitions for the flexibility measurement of optical subsystems.

Parameter	Definition
N	Number of BVTs in a subsystem/number of input ports on the WSS/SSS/MCS

M	Number of output ports on the WSS/SSS/MCS
c	Number of BVTs transmitting at a time/number of active input ports on the WSS/SSS/MCS
$(N-c)$	Number of BVTs that are off/number of input ports of the WSS/SSS/MCS that are not active
W	Number of tunable wavelength channels
k	Spectral granularity factor
kW	Number of tunable spectral slots
D	Total number of optical carriers
u	Number of optical carriers transmitted at the same time
$D-u$	Number of optical carriers that are off
B	Number of programmable modulations formats
E	Number of programmable symbol rates
r	Number of programmable electrical subcarriers
a	Number of copies of the same colour of wavelength channel/spectral slot that can be passed from c active input ports to M output ports at the same time.
$(c-a)$	Number of copies of the same colour of wavelength channel/spectral that are blocked on c active input ports
x	Maximum number of copies of the same colour of wavelength channel/spectral slot that can be passed from c active input ports to M output ports at the same time.
i	Number of active input ports on the MCS that are unblocked
$(c-i)$	Number of active input ports on the MCS that are blocked.

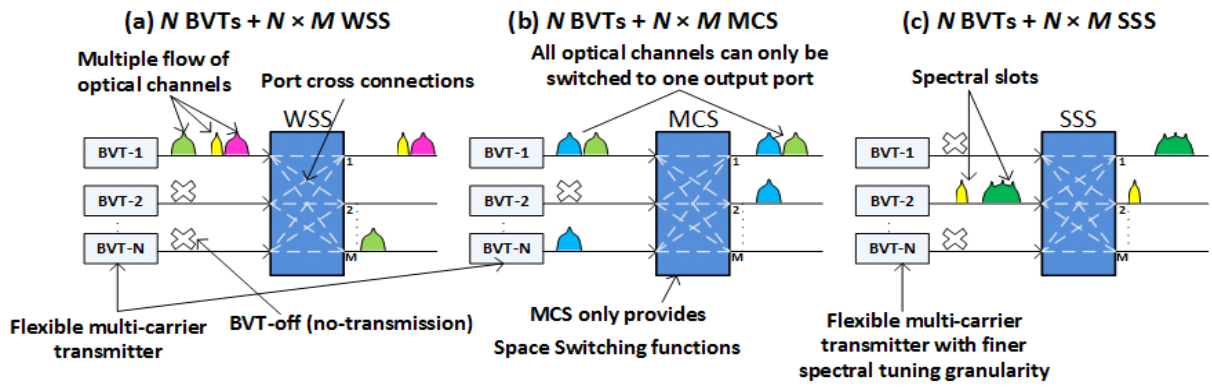

Figure 4.7: Different optical transmission and switching subsystems.

4.4.1 N BVTs + $N \times M$ WSS

BVTs only offers channel flexibility. Nevertheless, when a BVT is combined with an $N \times M$ WSS to form a subsystem, spectrum and space switching flexibility is achieved in addition to channel flexibility. Figure 4.7(a) illustrates a subsystem which consist of N BVTs combined with an $N \times M$ WSS. Each BVT is a flexible multi-carrier transmitter with programmable features for fixed grid networks. Thus, a single or multiple flows optical channels of different/same bandwidth can be fed into the input ports of the WSS which can either be blocked or switched to the same or different output ports. This subsystem design has application in data centres networks, and provides CDC functionality for ROADMs provided the lasers are tunable and the WSS design is without wavelength contention. To obtain the flexibility, the following assumptions are made: the subsystem is considered as a single system, each transponder is a multicarrier BVT with programmable features, the N number of BVTs are synchronized (i.e., if more than one BVT is transmitting, the same number of carriers and the same colours of wavelength channels are transmitted at the same time across all transmitting BVTs), and finally, the $N \times M$ WSS is without contention. The steps mentioned below show the derivation process to obtain the flexibility measurement model.

- Suppose c is the number of BVTs transmitting at the same time and the number of active input ports on the $N \times M$ WSS (i.e., if two BVT are transmitting, two of the N input ports of the WSS are active), and $(N - c)$ is the number of BVTs that are off and the number of input ports of the WSS that are inactive. The different number of ways $(N - c)$ can be chosen from N is

$$\binom{N}{N-c} \tag{4.17}$$

- Suppose D is the number of optical carriers on each of the active BVTs, u is the number of optical carriers that are transmitted at the same time, and $(D-u)$ is the number of optical carriers that are off. The different number of possible ways that $(D-u)$ can selected from D optical carriers is

$$\binom{D}{D-u} \tag{4.18}$$

- For c active BVTs, suppose r is the number of electrical subcarriers, B is the number of programmable modulation formats, E is the number of programmable symbol rates, and W is the number of tunable wavelength channels. The different number of ways that r electrical carriers can be independently configured to B modulations, E symbol rates, and injected unto u optical carriers with different wavelength channels is

$$\left(\left(\sum_{s=1}^{r}(BE)^r\right)^c\right)^u \frac{W!}{(W-u)!} \tag{4.19}$$

- For the c active ports on the $N \times M$ WSS, suppose a is number of copies of the same colour of wavelength channels that can be passed from c active input ports to M output ports at the same time, $(c-a)$ is the number of copies of the same colour of wavelength channels that are blocked on c active input ports, and x is the maximum number of copies of the same colour of wavelength channels that can be passed from c active input ports to M output ports at the same time. The different number of ways that $(c-a)$ can be chosen from c active input ports is $\binom{c}{c-a}$, and the number of different ways that a can be switched to M output ports without wavelength collisions is $\frac{M!}{(M-a)!}$. Thus, for all possible values of a, the number of possible states is

$$\left(\sum_{a=0}^{x}\frac{M!}{(M-a)!}\binom{c}{c-a}\right) \tag{4.20}$$

Combining Equation.(4.17), Equation.(4.18), Equation.(4.19) and Equation.(4.20), and summing all possible values of c, u and when the subsystem is off, the flexibility of the subsystem is

$$F(S) = log\left(\sum_{c=1}^{N}\left(\sum_{u=1}^{D}\left(\left(\sum_{r=1}^{r}(BE)^{r}\right)^{c}\right)^{u}\frac{W!}{(W-u)!}\binom{D}{D-u} \times \left(\sum_{a=0}^{x}\frac{M!}{(M-a)!}\binom{c}{c-a}\right)^{u}\binom{N}{N-c}\right)+1\right) \quad (4.21)$$

$$W \geq D \; and \; if \; c \leq N \; x = c, elseif \; c > M, x = M$$

This model can be used to measure any configuration of the N BVTs + $N \times M$ WSS. Note that Equation. (4.21) can only be applied when the number of tunable wavelength channels W is equal or greater than the number of optical carriers D ($W \geq D$). In addition, when $c \leq M$, $x = c$, this is because the maximum number of wavelength channels of the same colour that can be successfully passed at the same time from c active input ports to M output ports without wavelength collisions at M output ports is c. Otherwise, when $c > M, x = M$, this is because the maximum of number of wavelength channels of the same colour that can be successfully passed at the same time from c active input ports to M output ports without wavelength collisions at M output ports is M.

4.4.2 N BVTs + $N \times M$ Multicast switch

Figure 4.7(b) illustrates the N BVTs + $N \times M$ MCS subsystem design. Each BVT is a flexible multi-carrier transmitter and therefore provides channel flexibility for fixed grid networks. Because the MCS does not provide filtering functions, this subsystem provides only space switching and channel flexibility. Therefore, super channels or multiple channel flows generated from a BVT can only be switched to one output port and cannot be sliced across different output ports therefore providing low connectivity. This subsystem design provides CDC functionalities in ROADMs. The flexibility of this subsystem is modelled in a similar way to the N BVTs + $N \times M$ WSS but without spectrum switching flexibility. The definition of parameters for the BVT is the same as the parameters of the BVT described in the N BVTs + $N \times M$ WSS subsystem. However, for the MCS, c indicates the number of input ports that are active, i indicates the number of active input ports that are unblocked (i.e., allow optical

channels to pass at a time), and $(c-i)$ is equal to the number of active input ports that are blocked. Thus, flexibility of the subsystem is

$$F(S) = log\left(\sum_{c=1}^{N}\left(\begin{array}{c}\sum_{u=1}^{D}\left(\left(\sum_{r=1}^{r}(BE)^r\right)^c\right)^u \frac{W!}{(W-u)!}\binom{D}{D-u} \\ \times\left(\sum_{i=0}^{x}\frac{M!}{(M-i)!}\binom{c}{c-i}\right)\binom{N}{N-c}\end{array}\right)+1\right) \tag{4.22}$$

$$W \geq D \; and \; if \; c \leq N \; x = c, elseif \; c > M, x = M$$

4.4.3 *N* BVTs + *N* × *M* SSS

Figure 4.7(c) illustrates the N BVTs + $N \times M$ SSS subsystem design. The BVTs are flexible multi-carrier transmitters equipped with tunable lasers with finer tuning spectral granularity for flexible grid networks. This subsystem provides channel flexibility, space switching flexibility and spectrum switching flexibility with a finer spectral granularity when compared to the N BVTs + $N \times M$ WSS subsystem. This subsystem design is vital for EONs and realization of flexible grid ROADM with CDC features. The flexibility of this subsystem is derived in a similar way to the N BVTs + $N \times M$ WSS subsystem, the only difference is the finer spectral granularity of the BVTs and SSS. Suppose the spectral granularity factor is k and a single or multiple flows optical spectral slots can be fed from the BVT into the SSS, and can be blocked or switched by the SSS to the same or different output ports. The flexibility of the subsystem is

$$F(S) = log\left(\sum_{c=1}^{N}\left(\begin{array}{c}\sum_{u=1}^{D}\left(\left(\sum_{r=1}^{r}(BE)^r\right)^c\right)^u \frac{kW!}{(kW-u)!}\binom{D}{D-u} \\ \times\left(\sum_{a=0}^{x}\frac{M!}{(M-a)!}\binom{c}{c-a}\right)^u\binom{N}{N-c}\end{array}\right)+1\right) \tag{4.23}$$

$$kW \geq D \; and \; if \; c \leq N \; x = c, elseif c > M, x = M$$

4.5 Design Trade-offs of Optical Transmission and Switching Subsystems

To evaluate the impact of varying the number of optical carriers of BVTs on the flexibility of different subsystems, Figure 4.8 depicts the measured flexibility of the 4 BVTs + 4 × 16 WSS, 4 BVTs + 4 × 16 MCS and 4 BVTs + 4 × 16 SSS under the same design conditions and across

different number optical carriers per BVT. The flexibility is measured by varying the different number of optical carriers while other parameters are kept constant ($B = 5, E = 5\ and\ W = 30$). From Figure 4.8, it is noted that increasing the optical carriers has the greatest impact on the flexibility of the 4 BVTs + 4 × 16 SSS compared to the other subsystems due to finer spectral granularity of the SSS and BVTs. Furthermore, the increase in the number of optical carriers has a greater impact on the flexibility of the 4 BVTs + 4 × 16 WSS than the 4 BVTs + 4 × 16 MCS. This is attributed to the fact that the 4 BVTs + 4 × 16 WSS provides space and spectrum switching functions, i.e., slicing of optical channels across different output ports while the 4 BVTs + 4 × 16 MCS only provides space switching functions.

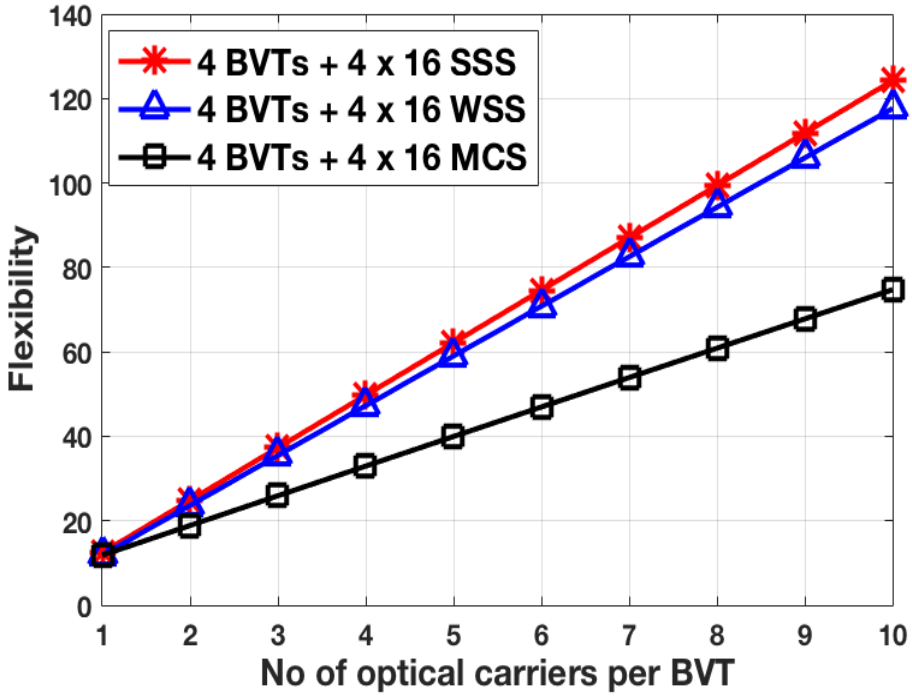

Figure 4.8: Comparison of the flexibility of 4 × 16 WSS/MSC/SSS across different numbers of optical carriers per BVT.

To evaluate the impact of varying the number of modulation formats of BVTs on the flexibility of different subsystems, Figure 4.9 shows the measured flexibility between the 4 BVTs + 4 × 16 WSS, 4 BVTs + 4 × 16 MCS and 4 BVTs + 4 × 16 SSS under the same design conditions and across different numbers of programmable modulations while other parameters are kept constant ($E = 5, D = 10\ and\ W = 30$). It is observed that there is a general increase in the flexibility of all the subsystems as the number of modulation formats increases. In more detail, the 4 BVTs + 4 × 16 SSS demonstrates the highest flexibility due to the spectral granularity of the SSS. The 4 BVTs + 4 × 16 WSS ranks second in flexibility and demonstrates a higher

flexibility than the 4 BVTs + 4 × 16 MCS because the WSS provides space and spectrum switching flexibility while the MCS supports only provides space switching flexibility. Furthermore, it is observed that the flexibility has a constant impact on all subsystems as the difference in flexibility between the subsystems remain constant as the number of modulation formats increases. It is also noted in Figure 4.9, that the pace of increase in flexibility reduces as the number of programmable modulation formats increases. Thus, comparing Figure 4.8 and Figure 4.9, it can be observed that for all subsystems, increasing the number of optical carriers has a greater impact on the pace of increase in flexibility than increasing the number of programmable modulation format.

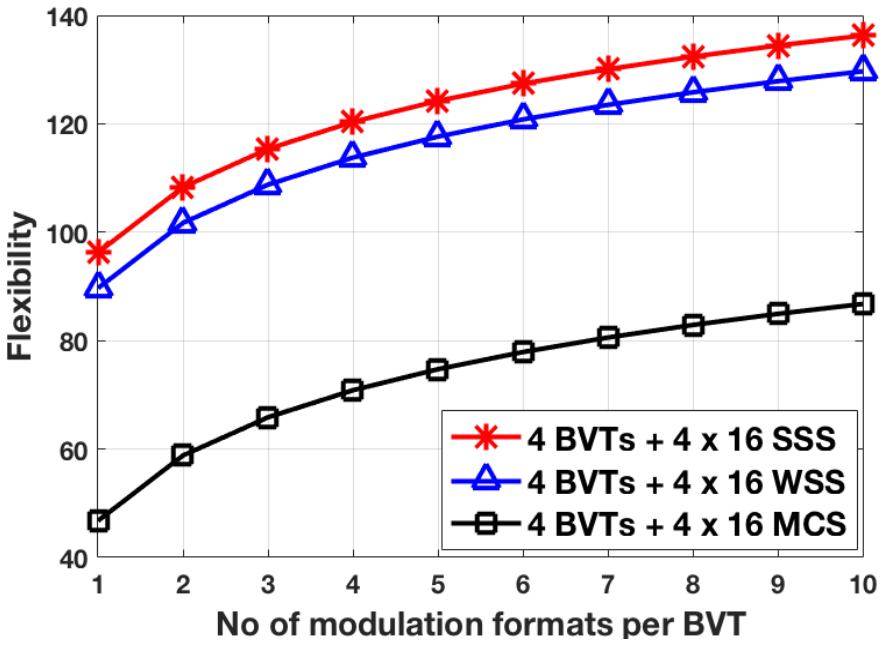

Figure 4.9: Comparison of the flexibility of 4 × 16 WSS/MSC/SSS across different numbers of programmable modulation formats.

In order to evaluate the design trade-off between the different subsystems described in the previous section, Table 4.6 presents different design configurations of optical transmission and switching subsystems while Figure 4.10 displays measured flexibility and other KPIs of each of the subsystems. The fibre connectivity of the subsystem is the number of output ports which can be connected to different fibres, and the optical carrier connectivity is the number of optical carriers in each BVT that can be sliced to different destinations. The maximum capacity is calculated with Equation.(4.24) where $MaxOC$ is the total number of optical carriers in a BVT,

SR_{max} is equal to the maximum achievable symbol rate, and SM_{max} is equal to the modulation format with the highest number of bits per symbol.

$$Maximum\ capacity = MaxOC \times SR_{max} \times log_2(SM_{max}) \quad (4.24)$$

The minimum granularity is the lowest achievable bitrate and is calculated with Equation.(4.25), where $MinOC$ is equal to 1 (i.e., 1 optical carrier), SR_{min} is equal to the minimum achievable symbol rate, and SM_{min} is equal to the modulation format with the lowest number of bits per symbol.

$$Minimum\ granularity = MinOC \times SR_{min} \times log_2(SM_{min}) \quad (4.25)$$

Table 4.6: Different subsystem configurations

Subsystem Config	No of carriers per BVT	Wavelength channels/ Spectral slots	Modulation formats	Symbol Rates (Gbaud)	Maximum Bitrate per (Gb/s)	Maximum Spectral efficiency	Minimum Granularity (Gb/s)
8 BVT + 8 ×12 WSS	8	80	BPSK, QPSK, 16QAM, 32QAM, 64QAM	30	1440	3.6	30
8 BVT + 8 × 24 MCS	10	80	BPSK, QPSK, 16QAM, 32QAM, 64QAM	14,18, 22, 26, 30	1800	3.6	14
4 BVT + 4 × 16 SSS	10	160(slots)	QPSK, 64QAM	14,18, 22, 26, 30	1800	4.8	28

*Required bandwidth for each optical channel is equal to symbol rate + 5GHz guard band

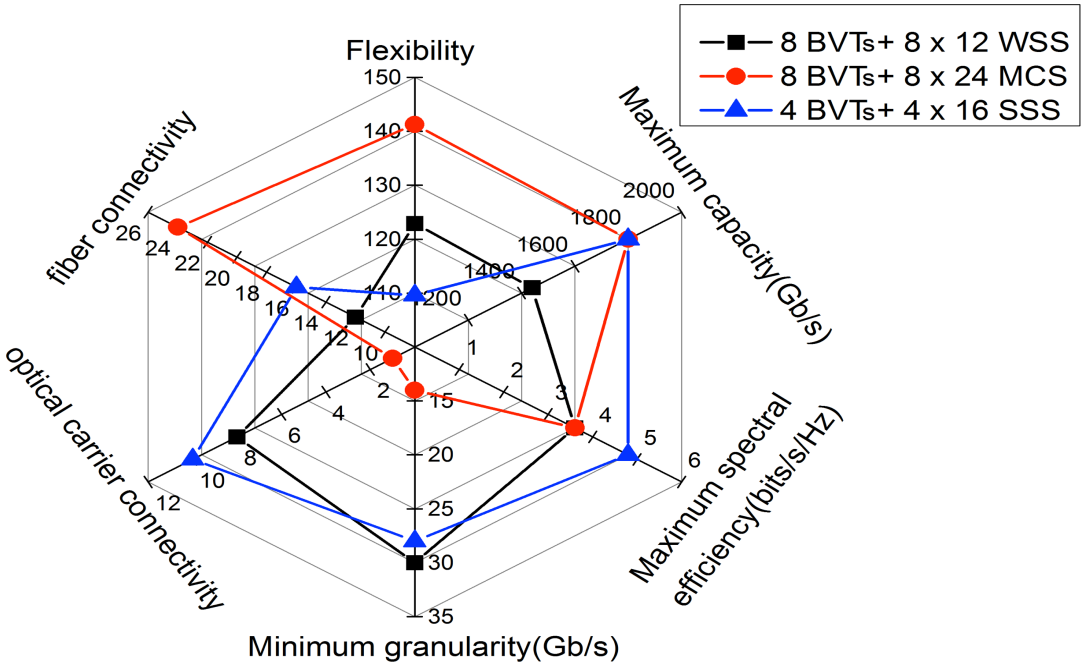

Figure 4.10: KPIs for the different subsystem configurations in Table 4.6.

The following observation in design trade-offs between KPIs of the different subsystems can be noted from Figure 4.10. First, it can be observed that the subsystem 4 BVTs + 4 × 16 SSS provides the lowest flexibility and has an equivalent maximum capacity to the 8 BVTs + 8 × 24 MCS, but at a higher spectral efficiency due to improved spectrum utilizations with a spectral slot slice of 12.5GHz. Secondly, the 8 BVTs + 8 × 24 MCS has the highest flexibility and fibre connectivity, however, it has the lowest optical carrier connectivity due to the fact that the MCS only provides space switching functions and cannot independently slice optical channels to different output ports. Finally, the 8 BVTs + 8 × 12 WSS has the second highest flexibility and has a higher optical carrier connectivity than 8 BVTs + 8 × 24 MCS but with a lower fibre connectivity. The evaluated design trade-offs highlight the relationship between multiple levels of flexibility and associated KPIs of different subsystem, and provides an insight into the selection of different combinations of optical components to build subsystems with different functionalities. For instance, in a scenario where a subsystem with high fibre connectivity, flexible channels, flexible rate and low optical carrier connectivity is required. For this scenario, the BVT design should contain tunable fixed grid lasers and programmable transmitter features. For the switching component, since optical carrier connectivity is negligible, an acceptable port configuration of MCS which supports only space switching

functions maybe used to build the subsystem as slicing of optical carriers to different destinations is not required.

4.6 Chapter Conclusion

This chapter presented models to measure the flexibility of a range of BVT configurations with programmable transmission features and different optical transmission and switching subsystems. The flexibility of different BVT modules were measured and compared with other measurable KPIs and design features. A comparative evaluation of results and theoretical analysis of various BVT modules showed that flexibility, connectivity and through loss have a proportional relationship, while no direct correlation between flexibility, capacity, spectral efficiency, cost and lightpath reach was noted. Also, results show that for BVTs, varying the number of optical carriers has a greater impact of flexibility than varying the number of programmable modulation formats. The flexibility of different optical transmission and switching subsystems under the same/different design conditions were also measured. Results showed under the same design condition, the BVTs and SSS subsystem demonstrated the highest flexibility, the BVTs and WSS subsystem ranks second, while the BVTs and MCS subsystem demonstrated the lowest flexibility. However, the case is not the same for the subsystems under different design conditions. Results showed that the combination of optical components with different design features can build subsystems which offer different levels of flexibility and functionalities.

5 Reconfigurable and Conventional Hybrid Disaggregated Data Centres

5.1 Introduction

In this chapter, the design features of the dRedBox architecture, conventional hybrid disaggregated DC architectures, and different arrangement approaches for disaggregated resource pools are illustrated and described. Furthermore, an illustration and description of the implementation of VM network requests across the various hybrid disaggregated DC architectures is presented. Finally, the structure and working principle of algorithms and network strategies to select EPS/OCS services, create multilayer custom topologies, and allocate network resources to build VM network requests on the various disaggregated DC architectures are presented.

5.2 Features of Hybrid Disaggregated Data Centre Architectures

This section introduces the various hybrid disaggregated DCs which are examined in this thesis. A cluster of dRedBox DC architecture is displayed in Figure 5.1(a). Each tray is composed of CPU bricks which hosts an MPSOC that embeds multiple cores, and memory bricks. Both the CPU and memory bricks are embedded with a hybrid and programmable electronic packet/circuit switch and optoelectronic mid-board with optical transceivers. The hybrid and programmable electronic packet/circuit switch can dynamically deploy electronic packet or circuit switching services to any I/O ports to handle different networking requirements. This allows for programmable ratios of packet to circuit switched services at runtime. The edge of tray (EoT), ToR, and top of cluster (ToC) are a hierarchical layer of optical circuit switches which provide tray, rack, and cluster level optical networking. Figure 5.1(b) and Figure 5.1(c) present conventional hybrid EPS/OCS architectures that have a fixed ratio of packet to circuit services. Figure 5.1(b) presents a cluster of 1-Tier hybrid disaggregated DC (1-Tier-H) architecture, and Figure 5.1(c) presents a cluster of 3-Tier hybrid disaggregated DC (3-Tier-H) architecture. The 1-Tier-H tray is composed of CPU bricks which hosts MPSOC that embed multiple CPU cores and memory bricks. Both CPU and memory bricks are embedded with optoelectronic mid-boards with optical transceivers and do not have any programmable electronic packet/circuit switches. EPS is carried out on dedicated EoT electronic packet switches, and the network outside tray is supported by an optical network.

The 3-Tier-H tray is similar to the 1-Tier-H tray. However, EPS is supported by dedicated ports on all network tiers (EoT, ToR, and ToC), while the remaining ports are supported by an optical network with a hierarchical layer of optical circuit switches (EoT, ToR, and ToC) that provide optical networking.

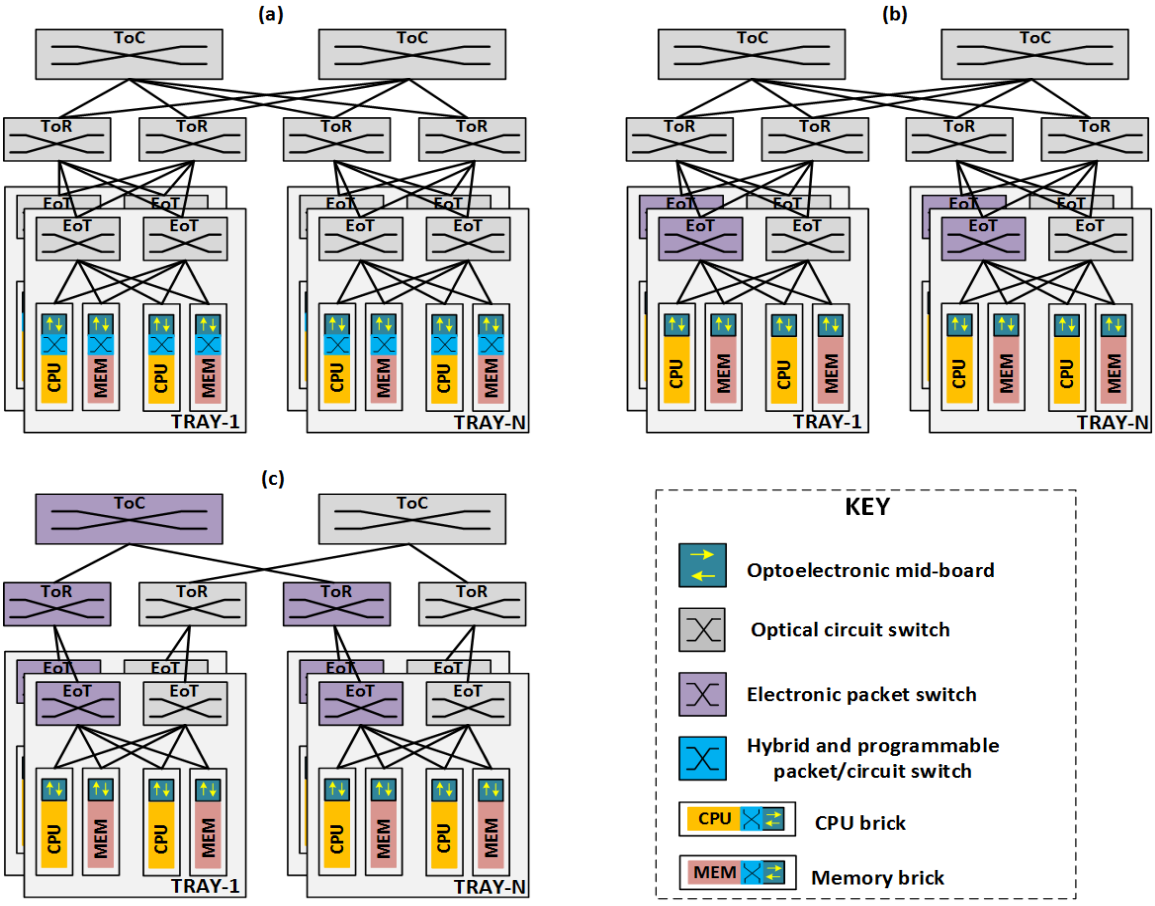

Figure 5.1: Hybrid disaggregated data centre architectures. (a) dRedBox. (b) 1-Tier-H. (c) 3-Tier-H.

Different configurations of the presented DC architectures can be achieved by varying the numbers of EoTs per tray, ports/transceivers per brick, trays in a rack, port configurations of optical switches, and the arrangement of CPU and memory brick resource pools. There are three different arrangement approaches for disaggregated resource pools in DCs which include: heterogeneous racks with stacks of heterogeneous trays category (RDC-1), heterogeneous racks with stacks of homogeneous trays category (RDC-2), and homogeneous racks with stacks of homogeneous trays category (RDC-3). A heterogeneous tray has both CPU bricks and memory bricks embedded on the same tray while a homogenous tray has CPU only or memory only bricks embedded on a single tray. The RDC-1 is illustrated in Figure 5.1, it can be

observed that each rack contains stacks of heterogeneous trays. In the RDC-2, each rack consists of a mixture of CPU only trays and memory only trays. The concept of the RDC-2 is illustrated in Figure 5.2.

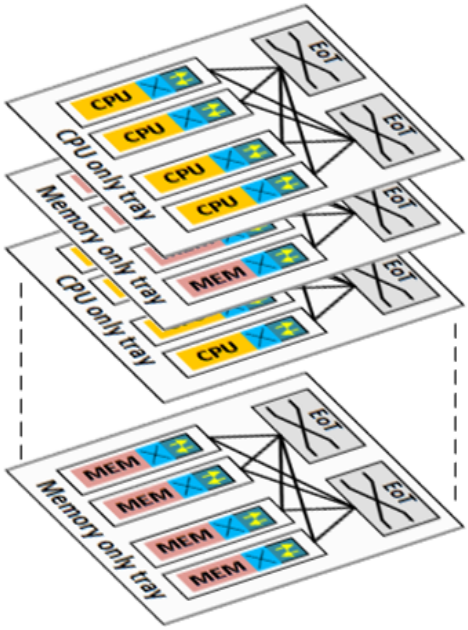

Figure 5.2: Heterogeneous racks with stacks of homogeneous trays.

In the RDC-3, each rack houses only one single resource type, i.e., each rack consists of stacks of either CPU only trays or memory only trays. The concept of the RDC-3 is illustrated in Figure 5.3.

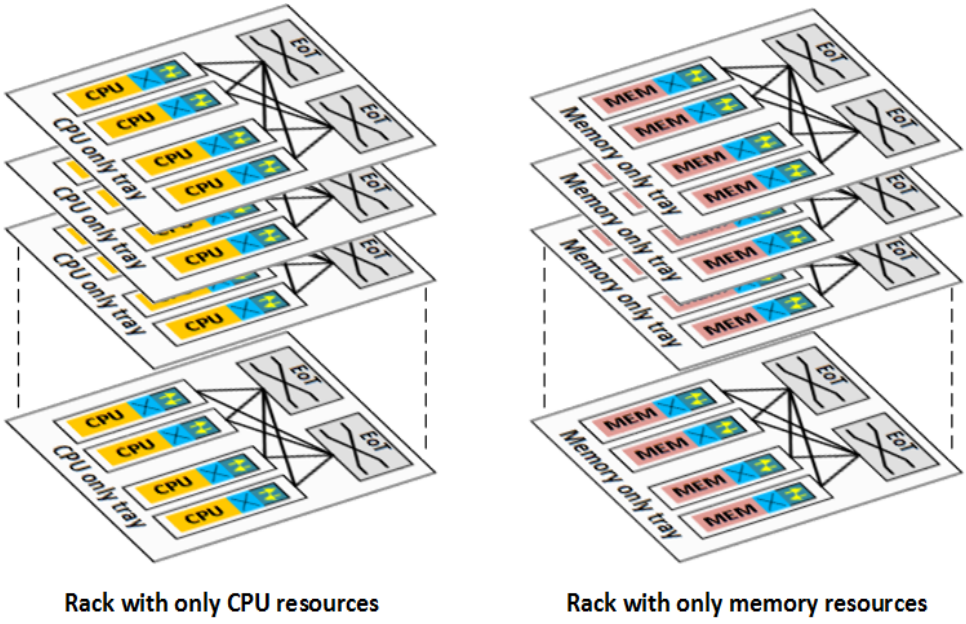

Figure 5.3: Homogeneous racks with stacks of homogeneous trays.

5.3 Implementation of VM network requests in Hybrid Disaggregated Data Centres

To understand the working principle of the previously described DC architectures, some examples of building VM network requests on the disaggregated DC architectures are highlighted. In this thesis, it is assumed that the IT resource requirements for VMs have already been allocated, and the requests considered are network requests to build VMs on disaggregated resources (and not associated with the network bandwidth requests between VMs or a VM and the cloud or user). Figure 5.4(a) and Figure 5.4(b) present two VM network request matrices that are to be deployed in the dRedBox and conventional tray architectures displayed in Figure 5.1. The transceivers embedded on the bricks and electronic packet switches are 10G transceivers, and all the links in the DC network are bidirectional and support a capacity of 10Gb/s in each direction. The rows and columns of the VM network request matrices in Figure 5.4 represents the CPU bricks and memory bricks, respectively. Each element in the matrix represents the required bandwidth for a single directional network link to be established between a CPU brick and memory brick (i.e., the total bandwidth requirement for a complete roundtrip communication between a CPU and memory brick is 2 × the bandwidth requirement of a single directional transaction). The VM network request 1 in Figure 5.4(a) indicates that CPU brick 1 requires network links to memory brick 2 and memory brick 4 with bandwidth requirements of 3 Gb/s and 2 Gb/s, respectively, and CPU brick 3 requires network links to memory brick 2 and memory brick 4 with bandwidth requirements of 3 Gb/s and 1 Gb/s, respectively. The VM network requests can be served using an EPS logical/virtual topology, an OCS logical/virtual topology or a combination of both. Building an EPS logical/virtual topology and serving the VM network request using EPS over OCS services is highly desirable due the statistical multiplexing features that EPS technology provides. This in turn leads to the best utilization of network and IT resources. However, care should be taken to limit its use due to latency overhead.

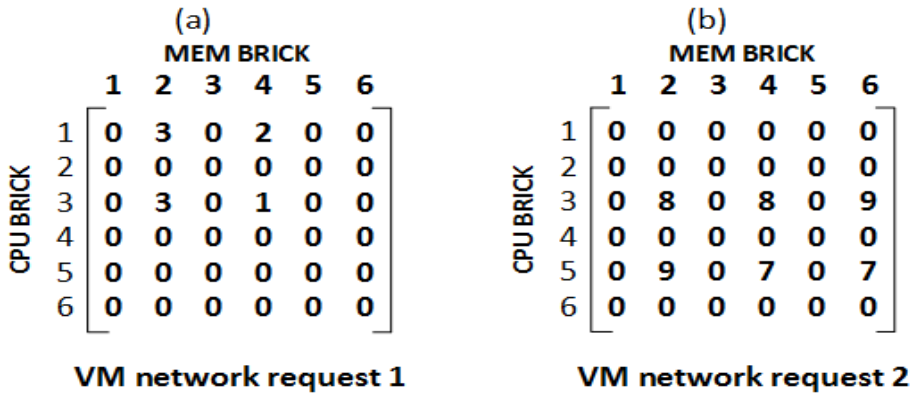

Figure 5.4: VM network request matrices.

Figure 5.5 displays EPS/OCS service allocation scenarios for deploying VM network request 1 and VM network request 2 on the dRedBox tray and the conventional tray architecture. It is assumed that VM network request 1 arrives before VM network request 2, and the resources attached to VM network request 1 are still in use when VM network request 2 arrives. For both scenarios, each brick on the dRedBox and conventional tray supports four I/O ports, i.e., four transceivers. Figure 5.5(a) illustrates the implementation of VM network request 1. When the bandwidth requirements between CPU and memory bricks of a VM network request are low (for example between 1 to 5 Gb/s), and each of the CPU bricks require network links to the same set of memory bricks, the multiple required network links can share network resources through statistical multiplexing, and an EPS virtual/logical topology can be built to implement the VM network request. Since VM network request 1 has low bandwidth requirements between CPU and memory bricks, and CPU bricks 1 and 3 require network links to memory bricks 2 and 4, an EPS virtual/logical topology is built, and EPS over OCS services are selected to implement VM network request 1 on the dRedBox and conventional tray architecture (details on the strategies for selecting EPS/OCS services are discussed in section 5.4.2). In the dRedBox tray, CPU brick 1 is configured as an electronic packet switch to perform statistical multiplexing of flows from CPU brick 1 and 3 to memory brick 2 and 4, while in the conventional tray architecture, the statistical multiplexing of flows from CPU brick 1 and 3 to memory brick 2 and 4 occurs in the EoT electronic packet switch.

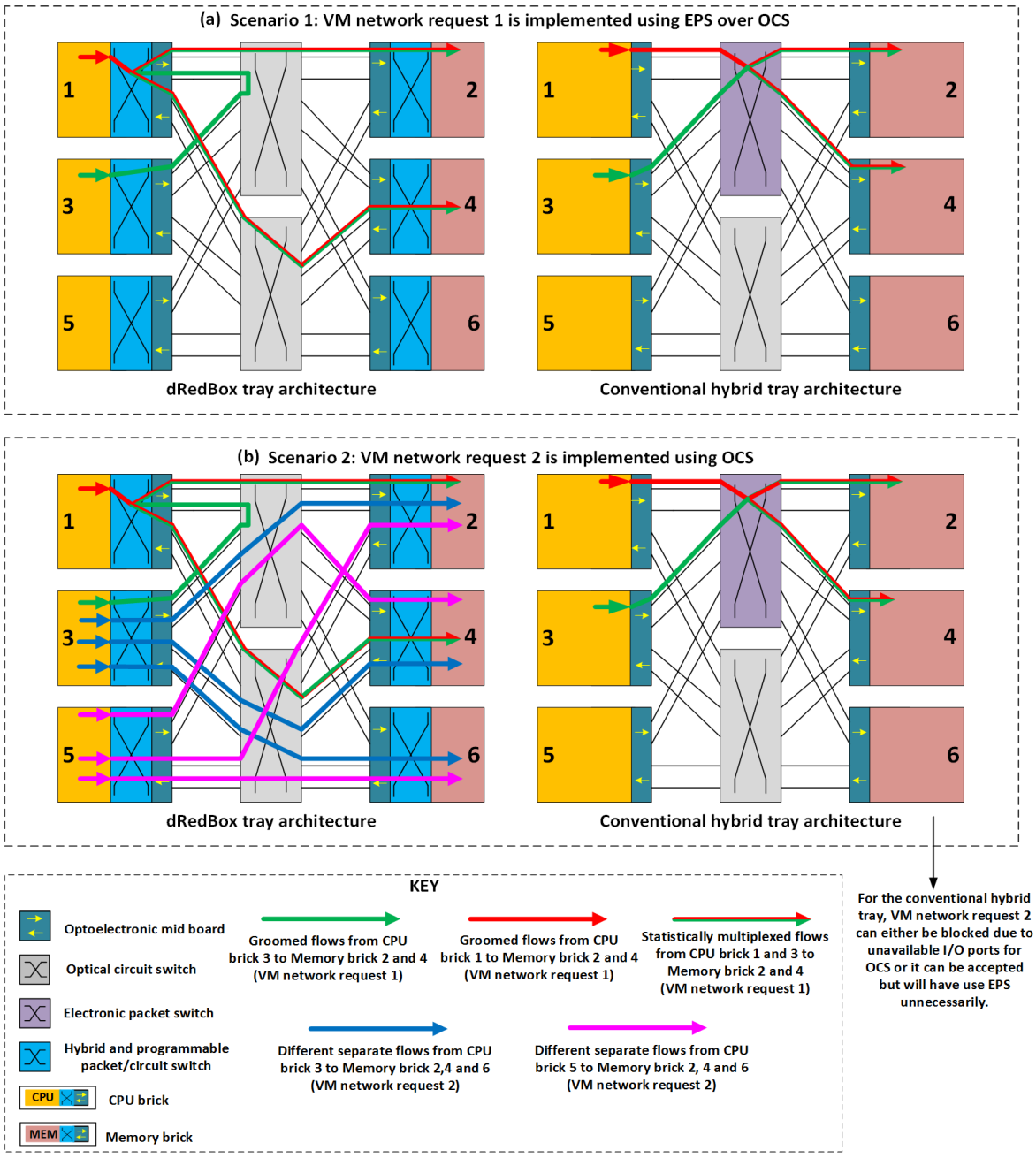

Figure 5.5: Deployment of VM network request using EPS/OCS services.

Comparing the two tray architectures in scenario 1 (Figure 5.5(a)), the dRedBox architecture (Figure 5.5(a) left) uses six brick I/O ports attached to six transceivers and twelve optical switch ports to implement the VM network request, whereas the conventional tray architecture (Figure 5.5(a) right) uses four brick I/O ports attached to four transceivers and eight electronic switch ports attached to four transceivers (i.e., the conventional tray architecture uses a total of eight transceivers). In terms of the number of transceivers, the dRedBox architecture offers transceiver savings when compared to the conventional architecture. In terms of the number of switch (network) ports and brick I/O ports, the dRedBox architecture uses more switch ports

and brick I/O ports than the conventional architecture. However, the dRedBox architecture is more cost-effective because network transmission is carried out over an optical switch unlike the conventional architecture, which uses a power-hungry electronic packet switch. Also, the dRedBox architecture also has the advantage of network function programmability and can deploy packet/circuit service to any of the I/O ports. Thus, either brick 1, 2, 3, or 4 can be configured to perform statistical multiplexing functions to merge flows from CPU brick 1 and CPU brick 3 to memory brick 2 and memory brick 4.

Figure 5.5(b) illustrates the implementation of VM network request 2. The bandwidth requirements between CPU and memory bricks for VM network request 2 are high (between 6 to 10 Gb/s), and multiple links cannot share network resources. Therefore, an OCS virtual/logical topology is built, and OCS services are selected to implement VM network request 2. In the dRedBox architecture, a point to point link between each pair of CPU brick and memory brick is established. Thus, three I/O ports in CPU bricks 3 and 5 and two I/O ports in memory bricks 2, 4 and 6 are selected, and the network optical switches are configured to implement the VM network request. In the conventional tray architecture, the VM network request 2 is either blocked due to unavailable I/O ports on the CPU bricks to support OCS or accepted but will have to use EPS resources unnecessarily.

5.4 Simulator Structure and Working Principle of the Proposed Algorithms

A simulator in MATLAB consisting of several algorithms was developed to investigate the performance of networking strategies and network switching services to deploy VM network requests on the dRedBox, 1-Tier-H and 3-Tier-H architectures. The simulator builds custom multi-layer virtual/logical topologies and allocates network resources at run-time to serve the VM network requests. The simulator is divided into three different stages as illustrated in Figure 5.6. The operations of the three different stages are consolidated in algorithm 1, and Table 5.1 displays the symbols and description for algorithm 1. In addition, the MATLAB function files in [117]–[120] from the MATLAB exchange file were used/modified in some sub-sections of the simulator.

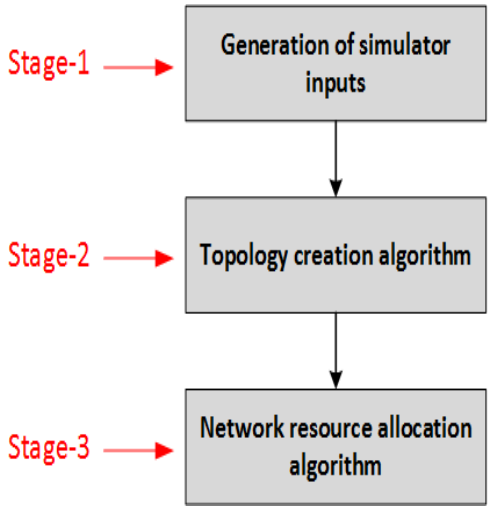

Figure 5.6: Simplified framework for simulator.

Table 5.1: Symbols and description for algorithm 1

Symbol	Description
$G(N, E)$	Graph representing a transparent DC network topology where N is the set of network nodes (Bricks, EoTs, ToRs and ToCs), and E is the set of physical links.
R	Set of sets where each set R_i represents a VM network request to be deployed in the DC network.
$N_{R_i}^{CPU}$	Set of CPU bricks in R_i.
$N_{R_i}^{MEM}$	Set of memory bricks in R_i.
B	VM network request matrix where each element is a bandwidth requirement for a network link between a CPU brick in $N_{R_i}^{CPU}$ and a memory brick in $N_{R_i}^{MEM}$. The rows and columns of the VM network request matrix represents the CPU bricks and memory bricks respectively.
P	Set of subsets where each subset P_i is a unique set or multiset of two elements and represents a possible way that different numbers CPU and memory bricks in a VM network request can be paired to build EPS virtual/logical topologies. The first element $p_{i,1}$ and second element $p_{i,2}$ of each subset represents the number of CPU bricks and memory bricks, respectively.
C	Set of subsets where each subset C_i contains a possible combination of the CPU bricks from $N_{R_i}^{CPU}$, and each element $c_{i,j}$ represents the j^{th} CPU brick in C_i.

CM	Set of memory bricks where element represents a memory brick which requires a network link from all the CPU bricks in C_i.
M	Set of subsets where each subset M_i contains a possible combination of the memory bricks from CM, and each element $m_{i,j}$ represents the j^{th} memory brick in M_i
V	A set which is the union all elements in C_i and M_i.
$TXcap$	Capacity of a transceiver.
$tb_{c_{i,j}}$	Sum of required bandwidth of network links from a CPU brick in C_i to all memory bricks in M_i.
$tb_{m_{i,j}}$	Sum of required bandwidth of network links from all CPU bricks in C_i to a memory brick in M_i.
$x_{c_{i,j}}$	Equals 1 if $tb_{c_{i,j}} \leq TXcap$, otherwise equals 0.
$x_{m_{i,j}}$	Equals 1 if $tb_{m_{i,j}} \leq TXcap$, otherwise equals 0.
K	Set which contains candidate bricks or electronic packet switches which can be used to build EPS virtual/logical topologies.
L	Number of bricks or electronic packet switches in K.
TDB	Database for EPS virtual/logical topologies.
OT	Set which contains the remaining network links in B that have not been used to create an EPS virtual/logical topology.
Z	Set of sets where each subset Z_s represents a created topology for the VM network request R_i, and element $z_{s,d}$ represents the d^{th} link in Z_s.
NZ	Number of generated topologies in Z.
f_{Z_s}	Number of links to established in Z_s.
$b_{z_{s,d}}$	Required bandwidth for a link $z_{s,d}$.
$eb_{z_{s,d}}$	Bandwidth in an existing network path for link $z_{s,d}$.
$h_{z_{s,d}}$	Equals to 1 if resources are available for link $z_{s,d}$, otherwise equals 0.

5.4.1 Simulator Inputs

The task achieved in the first stage of the algorithm is the generation of the simulator inputs and resource databases. Line-1 of the pseudocode (Algorithm 1) presents the simulator inputs which include: $G(N, E)$, R and P. In particular, P is a set of subsets, where each subset P_{i} is a unique set or multiset of two elements which contains a possible way that different numbers of CPU and memory bricks in a VM network request can be paired to build EPS virtual/logical topologies. The first element $p_{i,1}$ and second element $p_{i,2}$ of each subset represents the number of CPU bricks and memory bricks respectively. An example of the set P is illustrated in Figure 5.7. Assuming that the number of CPU bricks and memory bricks that can be allocated to a VM network request is between 2 to 4, P is generated with two steps. In the first step, the permutation with repetition of two-elements subsets from the set {2,3,4} is calculated. The set {2,3,4} is the range between the lowest number of CPU or memory bricks to the highest number of CPU or memory bricks that can be allocated in a VM network request which in this example is 2 and 4. In the second step, the generated two-elements subsets are sorted and rearranged starting from the subset with the elements that have the highest product to the subset with elements that have the lowest product. This is because building EPS virtual/logical topologies with higher number of CPU and memory brick combinations is beneficial, i.e., the higher the number of CPU and memory bricks that are paired to build EPS virtual/logical topologies, the more the number of network links in a VM network request can share network resources. This in turn leads to conservation of network resources for future use.

$$\{\{4,4\},\{4,3\},\{3,4\},\{4,2\},\{2,4\},\{3,2\},\{2,3\},\{2,2\}\}$$

Figure 5.7: The set P.

Algorithm 1

1: Inputs: $G = (N, E), R, P$
2: **for** each VM network request in R
3: Generate $N_{R_i}^{CPU}$, $N_{R_i}^{MEM}$, B
4: **for** each subset of P
5: Calculate $C = \binom{N_{R_i}^{CPU}}{p_{i,1}}$
6: **if** $C \neq \emptyset$
7: **for** each subset in C
8: Find CM
9: **if** the cardinality of $CM \geq p_{i,2}$
10: Calculate $M = \binom{CM}{p_{i,2}}$
11: **for** each subset in M
12: **if** $M_i \neq \emptyset$
13: **for** each CPU brick in C_i
14: Calculate $tb_{c_{i,j}}$
15: **if** $tb_{c_{i,j}} \leq TXcap$
16: $x_{c_{i,j}} = 1$
17: **else if**
18: $x_{c_{i,j}} = 0$
19: **end if**
20: **end for**
21: **if** $\sum_{c_{i,j} \in C_i} x_{c_{i,j}} = p_{i,1}$
22: **for** each memory brick in M_i
23: Calculate $tb_{m_{i,j}}$
24: **if** $tb_{m_{i,j}} \leq TXcap$
25: $x_{m_{i,j}} = 1$
26: **else if**
27: $x_{m_{i,j}} = 0$
28: **end if**
29: **end for**
30: **if** $\sum_{m_{i,j} \in M_i} x_{m_{i,j}} = p_{i,2}$
31: $V = C_i \cup M_i$

```
                    Run EPS placement strategy algorithm
                    Return Output: K
                    for each brick or electronic packet switch in K
                        Build EPS virtual/logical topology and store in TDB
                        Update B
                    end for
                end if
            end if
        end if
    end for
  end if
end for
if TDB is empty
  Store VM network request as an OCS virtual/logical topology in Z
else if TDB is not empty
  Compute OT
  if OT = ∅
    Store EPS virtual/logical topologies in Z
  else if OT ≠ ∅
    Combine OT with all created EPS virtual/logical topologies in TDB and
    store in Z
  end if
end if
for s = 1:NZ (for each topology in Z)
  for each link in the selected topology
    h_{z_{s,d}} = 0
    find already established paths
    if established paths exist
        for each established path
          if  b_{z_{s,d}} + eb_{z_{s,d}} ≤ TXcap
            h_{z_{s,d}} = 1
            break
```

```
65:                 end if
66:               end for
67:       else if established path does not exist & h_{z_{s,d}} = 0
68:           Find least congested path
69:           if resources are available on the found path
70:               h_{z_{s,d}} = 1
71:           end if
72:       end if
73:     end for
74:     if (∑_{z_{s,d}∈Z_s} h_{z_{s,d}}) = f_{Z_s}
75:         Accept and implement VM network request
76:     else if (∑_{z_{s,d}∈Z_s} h_{z_{s,d}}) ≠ f_{Z_s} & s < NZ
77:         Check next topology
78:     else if (∑_{z_{s,d}∈Z_s} h_{z_{s,d}}) ≠ f_{Z_s} & s = NZ
79:         VM network request is blocked
80:     end if
81:   end for
82: end for
```

5.4.2 Topology Creation Algorithm

The overall task achieved in the second stage of the algorithm is the creation of custom network topologies to implement the VM network request. The topologies could either be an EPS virtual/logical topology, an OCS virtual/logical topology or a combination of both. The first task of the algorithm is to search for suitable combinations of CPU and memory bricks that can be paired to build EPS virtual/logical topologies. When a VM network request is received in a DC, the first step generates $N_{R_i}^{CPU}$ a set of CPU bricks, $N_{R_i}^{MEM}$ a set of memory bricks, and B a VM network request matrix. Next, the set P is used to build sets of possible combinations of CPU bricks. For each subset of P, a set of subsets C is calculated by $\binom{N_{R_i}^{CPU}}{p_{i,1}}$, where each subset C_i contains a possible combination of $p_{i,1}$-CPU bricks. If the generated set C is not empty, for each subset C_i starting from the first, the algorithm searches for a set of memory bricks CM, where each of the memory bricks requires a network link from all the CPU bricks in C_i. If the cardinality of CM is greater than or equal to $p_{i,2}$, M, which is a set of subsets where each subset

M_i contains a possible combination of memory bricks from CM, is calculated by $\binom{CM}{p_{i,2}}$. In the event that C is empty, i.e., no CPU brick combination has been generated or the cardinality of CM is less than $p_{i,2}$, the CPU brick number $p_{i,1}$ is discarded, and next CPU brick number from the next subset of P is selected and used to generate another set of CPU brick combinations. Figure 5.8 illustrates an example of generated combinations of CPU and memory bricks with the different subsets of P from Figure 5.7, and from a VM network request with four CPU bricks {1,3,5,7} and four memory bricks {2,4,6,8} to be interconnected.

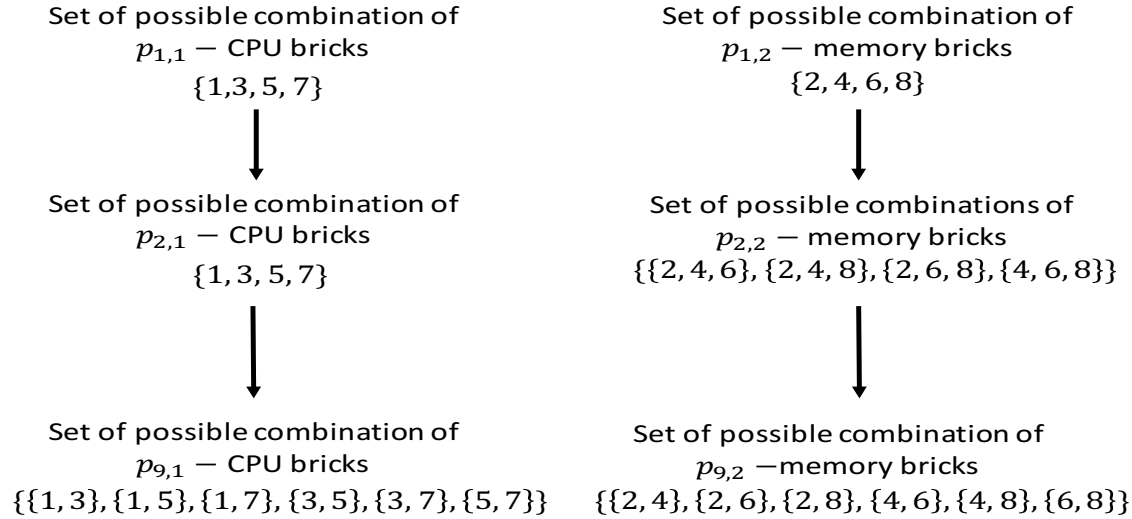

Figure 5.8: Example of generated combinations of CPU and memory bricks with the different subsets of P.

Afterwards, for each subset in M starting from the first one, if M_i is not empty, the algorithm calculates the $tb_{c_{i,j}}$, which is the sum of required bandwidth for network links from a CPU brick in C_i to all memory bricks in M_i. Then, if $tb_{c_{i,j}}$ from a CPU brick in C_i is less than or equal to the capacity of a transceiver $TXcap$, $x_{c_{i,j}}$ is marked as 1, otherwise marked as 0. This step is carried out for all the CPU bricks in C_i. If the summation of the variable $x_{c_{i,j}}$ (for all CPU bricks in C_i) is equal to $p_{i,1}$, the first condition to build an EPS virtual/logical topology is satisfied. After the first condition has been satisfied, the next line calculates $tb_{m_{i,j}}$, which is the sum of required bandwidth of network links from all CPU brick in C_i to a single memory brick in M_i. If the $tb_{m_{i,j}}$ is less than or equal to $TXcap$, $x_{m_{i,j}}$ is marked as 1, otherwise marked as 0. Then, if the summation of the variable $x_{m_{i,j}}$ (for all memory bricks in M_i) is equal to $p_{i,2}$, the second condition to build an EPS virtual/logical topology is satisfied. In the event that the summation of the variable $x_{c_{i,j}}$ (for all CPU bricks in C_i) is not equal to $p_{i,1}$ or the summation

of the variable $x_{m_{i,j}}$ (for all memory bricks in M_i) is not equal to $p_{i,2}$, the subset of memory bricks is discarded, and the next memory brick combination which is the next subset in M selected and checked.

Once the two conditions to build an EPS virtual/logical topology is satisfied, a set V, which is the union of CPU bricks in C_i and memory bricks in M_i, is generated. The next task of the algorithm is to run the EPS placement strategy algorithm for the selected combinations of CPU bricks and memory bricks in order to select the most suitable bricks or electronic packet switches to perform statistical multiplexing functions when creating an EPS virtual/logical topology. There are three different algorithms that can be selected at this stage: the congestion aware placement strategy algorithm, the network-hop aware placement strategy algorithm and random placement strategy algorithm. Figure 5.9 illustrates examples of different possible EPS virtual/logical topologies that are created from a CPU brick combination of {1,3} and a memory brick combination of {4,8} across the different disaggregated architectures. Figure 5.9(a), Figure 5.9(b), and Figure 5.9(c) presents a cluster of dRedBox, 1-Tier-H, and 3-Tier-H architectures, respectively. Each architecture consists of two racks which contain two trays each, and each tray contains a CPU and memory brick.

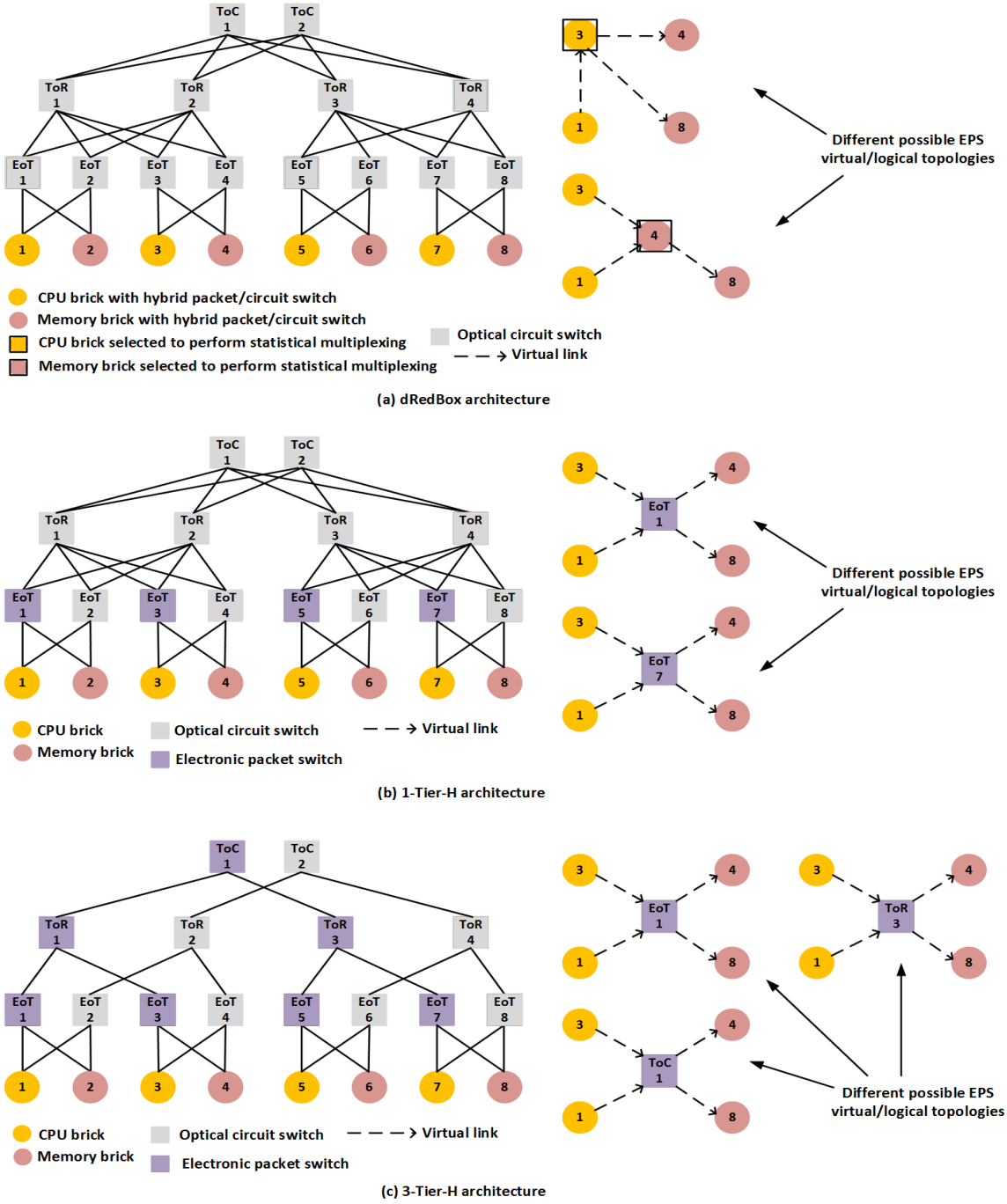

Figure 5.9: Examples of different EPS virtual/logical topologies. (a) dRedBox architecture (b) 1-Tier-H architecture. (c) 3-Tier-H architecture.

For the dRedBox architecture, in the congestion aware placement strategy algorithm, the brick with least number of utilized ports is given highest priority to be configured to perform statistical multiplexing functions in the created EPS virtual/logical topology. Algorithm 2 illustrates steps of the congestion aware strategy algorithm for the dRedBox architecture. For each brick in V, the number of utilized brick ports is calculated. Also, the number of network hops required for each of the bricks to perform statistical multiplexing of flows to and from

other bricks is calculated. Next, the bricks are sorted from the brick with the least utilized ports to the brick with the most utilized ports. In a situation where two or more bricks have the same number of utilized ports, those bricks are sorted from the brick which requires the least number of network hops to the brick which requires the most number of network hops. The sorted bricks are stored in K in an order that starts from the most suitable brick to least suitable brick. The network-hop aware placement strategy algorithm follows a similar procedure to the congestion aware placement strategy algorithm. The priority factor for creating an order of suitable bricks to perform statistical multiplexing is the number of required network hops. Thus, after the number of utilized brick ports and required number of network hops for each brick has been calculated, the bricks are sorted from the brick which requires the least number of network hops to the brick which requires the most number of network hops. In a scenario where two or more bricks have the same number of network hops, those bricks are sorted from the brick with the least utilized ports to brick with the most utilized ports. In the random placement strategy algorithm, the bricks are selected and sorted randomly without considering any factor. Figure 5.9(a) shows examples where possible EPS virtual/logical topologies are created with a CPU brick combination of {1,3} and a memory brick combination of {4,8} for dRedBox architecture. In one of the topologies, a CPU brick is selected to perform statistical multiplexing, while in the other topology, a memory brick is selected to perform statistical multiplexing.

Algorithm 2: Congestion aware placement strategy algorithm for dRedBox architecture

1: **for** each brick in V
2: Calculate the number of utilized brick ports
3: Calculate the number of network hops required when the brick is configured as an electronic packet switch
4: **end for**
5: Sort bricks from the brick with least utilized ports to brick with most utilized ports
6: Find any bricks with same number of utilized ports
7: **If** there are any bricks with the same number of utilized ports
8: Sort the selected bricks from the brick with the least number of network hops to the brick with most number of network hops
9: **end if**
10: Store sorted brick in K

For the 1-Tier-H architecture, the statistical multiplexing is performed on the EoT electronic packet switches and not on the bricks. The candidate EoT electronic packet switches that are considered to perform statistical multiplexing are the EoT switches which are attached to a tray, where either a CPU or memory brick in V is located. For instance, in Figure 5.9(b), the candidate EoT switches that will be considered to perform statistical multiplexing when building an ESP virtual/logical topology for a CPU brick combination of {1, 3} and a memory brick combination of {4, 8}, are EoT-1, EoT-3 and EoT-7. In the congestion aware algorithm, the number of utilized ports for each of the candidate EoT switches is calculated, and the number of network hops required when each of the candidate EoT switches is selected to perform statistical multiplexing of flows to and from the selected CPU and memory bricks is calculated. Next, the switches are sorted from the switch with the least utilized ports to the switch with the most utilized ports. In a situation where two or more switches have the same number of utilized ports, those switches are sorted from the switch that requires the least number of network hops to the switch that requires the most number of network hops. Next, the sorted EoT switches are stored in K in an order that starts from the most suitable switch to least suitable switch. In the network-hop aware placement strategy algorithm, the considered EoT electronic switches are sorted from the switch with the least required number of network hops to the switch with the most number of network hops. Also, in a situation where two or more EoT electronic packet switches have the same required number of network hops, those set of EoT switches are sorted from the EoT switch with the least utilized ports to the EoT switch with the most utilized ports. In the random placement strategy algorithm, the EoT switches are selected and sorted randomly without considering any factor.

For the 3-Tier-H architecture, statistical multiplexing can occur in EoT, ToR and ToC electronic packet switches. The candidate electronic packet switches that are considered to perform statistical multiplexing are the EoT switches that are attached to a tray where either a CPU or memory brick in V is located, the ToR switches that are attached to a rack where either a CPU or memory brick in V is located, and the ToC electronic packet switch of the cluster where either a CPU or memory brick in V is located. For instance, in Figure 5.9(c), the candidate electronic packet switches which will be considered to perform statistical multiplexing when building an ESP virtual/logical topology for a CPU brick combination of {1,3} and a memory brick combination of {4,8} are EoT-1, EoT-3, EoT-7, ToR-1, ToR-3 and ToC-1. The working principle of the congestion aware, network-hop aware, and random placement strategy algorithm follows a similar procedure to the 1-Tier-H architecture, the only

difference is that since electronic packet switches considered are on different networking layers and have different number of ports, the metric used for determining congestion is the percentage of utilized ports of each of the candidate switches instead of the number of ports.

For all the architectures, after the set K has been created, up to L number of EPS virtual/logical topologies can be created. Next, the EPS virtual/logical topologies that have been built are stored in TDB, and the VM network request matrix B is updated by removing all CPU to memory network links that have been used in the created EPS virtual/logical topology. Also, any subset in M that contains any memory brick that has been utilized in the previously built EPS virtual/ logical topology with the selected the CPU brick combination in C_i is discarded. After all the subset of P have been checked for all possible combinations of CPU and memory bricks to build EPS virtual/logical topologies, in the event that TDB is empty, i.e., no EPS virtual/logical topology has been built, an OCS virtual/logical topology for the VM network request is built and is stored in Z. Otherwise, if TDB is not empty, a set OT which contains any remaining links in B that have not been used to build an EPS virtual/logical topology is generated. If OT is not empty, the links in OT are combined with the EPS virtual/logical topologies in TDB to form hybrid EPS/OCS virtual/logical topologies which are stored in Z. In the event that OT is empty, i.e., there are no remaining links in the B to be established, all the EPS virtual/logical topologies in TDB are combined and stored in Z.

5.4.3 Network Resource Allocation

The next stage of the algorithm searches for network resources for the created topologies. After Z has been created, the algorithm searches for network resources for the first topology stored in Z. For each link to be established in the selected topology, the algorithm first searches to find an already established path with available bandwidth resources to perform optical grooming services. If there is an existing path with available bandwidth resources, i.e., $b_{z_{s,d}} + eb_{z_{s,d}} \leq TXcap$, the link is marked for optical grooming on the established path ($h_{z_{s,d}} = 1$). Alternatively, if there are existing paths but with no available bandwidth for optical grooming, or there are no existing paths for optical grooming, the algorithm then searches for free resources (I/O and network ports) by searching and selecting the least congested network path. If resources are available on the found network path, the link is marked to be deployed on the resources on that path ($h_{z_{s,d}} = 1$). If resources are available for all links, the VM network request is accepted. If resources are not available for any link, other stored topologies in Z are

checked. If resources are not available to any of the topologies, the VM network request is blocked.

5.5 Chapter Conclusion

In this chapter, the features of the different hybrid disaggregated DC architectures were presented. The DC architectures discussed include: dRedBox, 1Tier-H and 3-Tier-H. Also, three different arrangement approaches for disaggregated resource pools in DCs were presented. Next, an illustration of the implementation of VM network requests using EPS/OCS services on the dRedBox and conventional tray architectures was presented. Also, a performance trade-off in terms of number of utilized transceivers, brick I/O ports and switch ports of both tray architectures based on the illustration was reported. Finally, the simulator framework and working principle of the proposed algorithm were described. The simulator framework was broken down into three stages: simulator inputs, topology creation algorithm and network resource allocation. The task achieved in each of the stages were described. In particular, three different networking strategies in topology creation algorithm with objective of selecting the most suitable brick or electronic packet switch to perform statistical multiplexing functions when building an EPS virtual/logical topology to implement VM network requests were proposed and discussed.

The network strategies and algorithms presented in this chapter are key to evaluating the performance of different hybrid disaggregated DC architectures, different arrangement approaches of disaggregated resource pools, and providing insights into solutions for the challenges associated with the allocation of EPS/OCS services and network resources to build VMs on disaggregated resources.

6 Performance Evaluation of dRedBox Architecture

6.1 Introduction

In this Chapter, the performance of various networking strategies for deploying VM network requests across the dRedBox, 1-Tier-H, and 3-Tier-H architectures under different traffic patterns are evaluated in terms of blocking probability, number of successfully deployed VMs, network utilization, network capacity, and energy efficiency. Also, a cost evaluation of the different architectures is presented. Furthermore, the performance of the different arrangement approaches of CPU and memory resources in the dRedBox architecture is evaluated in terms of blocking probability, number of successfully deployed VMs, and number of utilized switch ports.

6.2 dRedBox Architecture versus Conventional Hybrid Disaggregated Data Centre Architectures

In this section, a comprehensive performance evaluation of the dRedBox, 1-Tier-H and 3-Tier-H architectures is presented and discussed.

6.2.1 Simulation Assumption

The configuration of the dRedBox, 1-Tier-H and 3-Tier-H architectures are simulated as illustrated in Figure 5.1. Each of the DC architectures consists of a cluster containing 4 racks with each rack consisting of 4 heterogeneous trays. Each tray is populated with 14 bricks, i.e., 7 CPU and 7 memory bricks. Each brick is interfaced with 24 10Gb/s transceivers. All the DC architectures have a 1:1 port connectivity subscription ratio between all networking layers and the same number of switch ports on all networking layers. Each of the EoT switches has 672 ports (i.e., the same performance of each EoT switch simulated can be achieved by connecting 8 switches with 84 ports each in parallel), each of the ToR switches has 2688 ports (i.e., the same performance of each ToR switch simulated can be achieved by connecting 12 switches with 224 ports each in parallel), and each of the ToC switches has 5376 ports (i.e., the same performance of each ToR switch simulated can be achieved by connecting 16 switches with 336 ports each in parallel).

There are different bandwidth requirements for local memory associated with the memory technology and processor technology. As mentioned in section 2.5.1, results from a previous

study and experiment in [82] showed that for certain applications, 20-40 Gb/s for remote memory access can achieve minimal (under 10%) application performance degradation. Furthermore, a related study in [83] reported that a bandwidth of 582 Mib/s (~ 5Gb/s) can be realized between a single CPU core and remote access memory. Also, the authors in [84] demonstrated remote memory access with 10 Gb/s links and up to 68% sustained memory bandwidth. In this thesis, VM network requests are randomly generated using the following parameters: 2 to 5 number of CPU bricks, 2 to 5 number of memory bricks, and bandwidth requirement between a CPU brick and a memory brick that varies between 1 to 5 Gb/s and 6 to 10 Gb/s, which is classified as low bandwidth traffic pattern (LBTP) and high bandwidth traffic pattern (HBTP), respectively (i.e., for a complete round trip transaction between one CPU and one memory brick, LBTP is 2 to 10 Gb/s and HBTP is 12 to 20 Gb/s). The authors in [17] used a similar bandwidth traffic pattern to evaluate the benefits of an OPS/OCS hybrid DC. Also, for each generated VM network request, each of the CPU brick requires a network link to each of the memory bricks. It is assumed that VM network requests arrive dynamically following a Poison process with a mean inter-arrival rate of 10 time units. A previous study on disaggregated DCs in [23] has used a similar approach. A holding time range of 525 to 2100 time units with increments of 525 time units was selected to ensure different levels of network load in the DC. Once the holding time of a successfully deployed VM network request expires, the resource attached to that VM network request are released.

A total of 400 VM network requests are generated, and the results obtained in this thesis are averaging 4 simulations runs for each point with a 95% confidence interval. The following abbreviations are used for the different EPS function placement strategies on the various architectures. The dRedBox congestion aware, dRedBox network-hop aware, and dRedBox random placement strategies are represented as D-COS, D-NES, and D-RAS, respectively. The 1-Tier-H congestion aware, 1-Tier-H network-hop aware, and 1-Tier-H random placement strategies are represented as 1-T-COS, 1-T-NES, and 1-T-RAS, respectively. Finally, 3-Tier-H congestion aware, 3-Tier-H network-hop aware, and 3-Tier-H random placement strategies are represented as 3-T-COS, 3-T-NES, and 3-T-RAS, respectively. The various EPS function placement strategies are evaluated in terms of blocking probability, network capacity, network utilization, energy efficiency, and cost. Blocking probability is the ratio of number of blocked requests to the total number of requests that have been processed in the DC network. Network capacity is a measure of the total number of bits per second transmitted on all the network links of DC network. Network utilization is the ratio of the number of utilized network resources

(number of used brick I/O ports and switch ports) to the total number of network resources (total number of brick I/O ports and switch ports in the DC network). Energy efficiency (J/b) is the ratio of power consumption (W) to network capacity (b/s). The parameters describing the cost and power consumption of components are set according to the values presented in Table 6.1. The cost of a 10G transceiver is assumed to be $50, and power per optical switch port is 0.05 W [121]. The remaining values in Table 6.1 are sourced from [14][19]. Also, the cost and power consumption are calculated based on the number of resources used.

Table 6.1: Power Consumption and Cost Values of DC network components

Component	Power (W)	Cost ($)
Optical switch port	0.05	500
Electronic switch port	12.5	500
Transceiver 10G	1	50

6.2.2 Random Bandwidth Traffic Variation

For the random bandwidth traffic variation simulation scenario, the VM network requests generated are randomly assigned to either LBTP or HBTP. Figure 6.1 shows the blocking probability of the different EPS placement strategies across the dRedBox, 1-Tier-H, and 3-Tier-H architectures at different holding times. The points plotted in Figure 6.1 represent the blocking probability after the last VM network request, i.e., 400th request, has been received and processed in the DC network. It can be noted from Figure 6.1 that the dRedBox architecture for all the EPS placement strategies has a lower blocking probability than the 1-Tier-H and 3-Tier-H architectures across the different holding times. Furthermore, despite the 3-Tier-H architecture having more levels for EPS than the 1-Tier-H architecture, Figure 6.1 shows that the 3-Tier-H and 1-Tier-H architecture display similar level of blocking probability across all EPS placement strategies and holding times. This is because both the 3-Tier-H and 1-Tier-H architecture have the same proportions of brick I/O ports dedicated to either optical circuit or electronic packet network. This limits resource availability to process different traffic variations (LBTH or HBTH). Thus, having different levels of EPS switches in the 3-Tier-H architecture has negligible impact on resource availability to serve VMs. In addition, it can be observed that D-COS performs the best (has the lowest blocking probability). At 2100 holding time units, D-COS demonstrates approximately 9% lower blocking probability than D-NES, 16% lower blocking probability than D-RAS, and 35% lower blocking probability than the 1-T-COS, 1-T-NES, 1-T-RAS, 3-T-COS, 3-T-NES, and 3-T-RAS.

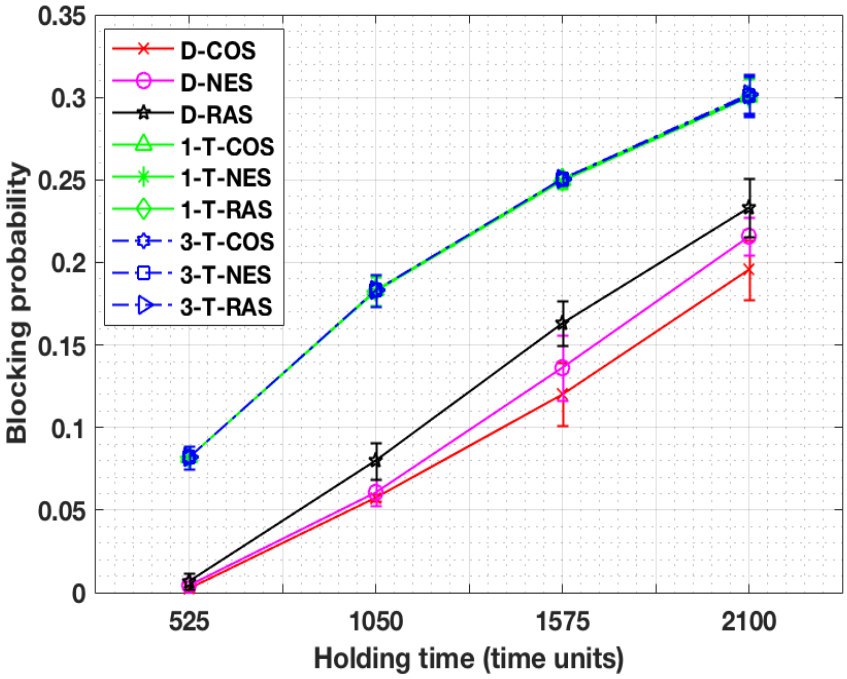

Figure 6.1: Blocking probability for random bandwidth traffic variation.

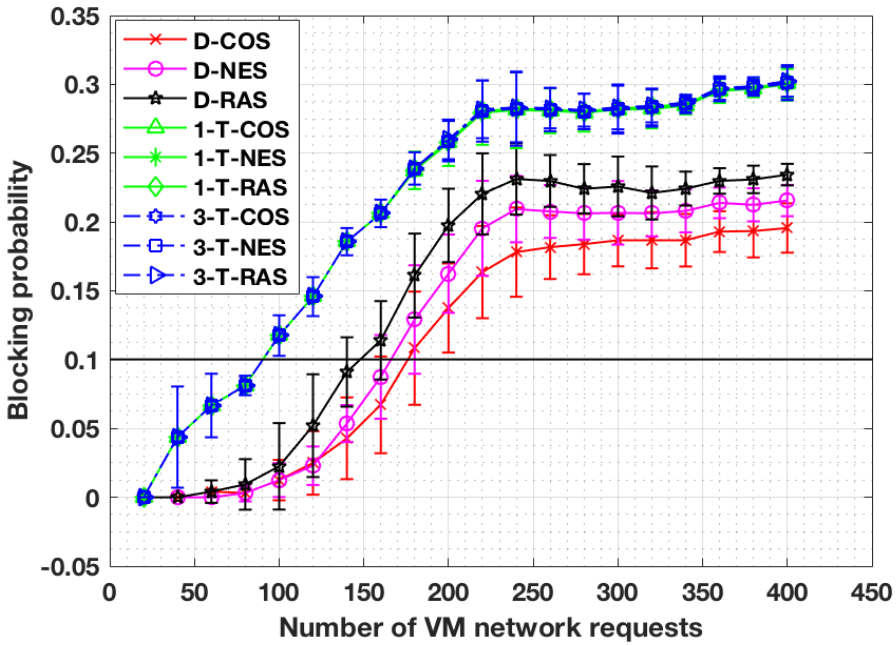

Figure 6.2: Blocking probability for 2100 holding time scenario.

Figure 6.2 shows the blocking probability of the 2100 holding time scenario with a line highlighting the 10% (0.1) blocking probability point. It can be observed in Figure 6.2 that the dRedBox architecture for all EPS placement strategies demonstrates lower blocking probability that the 1-Tier-H and 3-Tier-H architectures across different numbers of VM

network requests. The 0.1 blocking probability point is selected as the maximum threshold value to further evaluate the performance of the different EPS placement strategies and DC architectures.

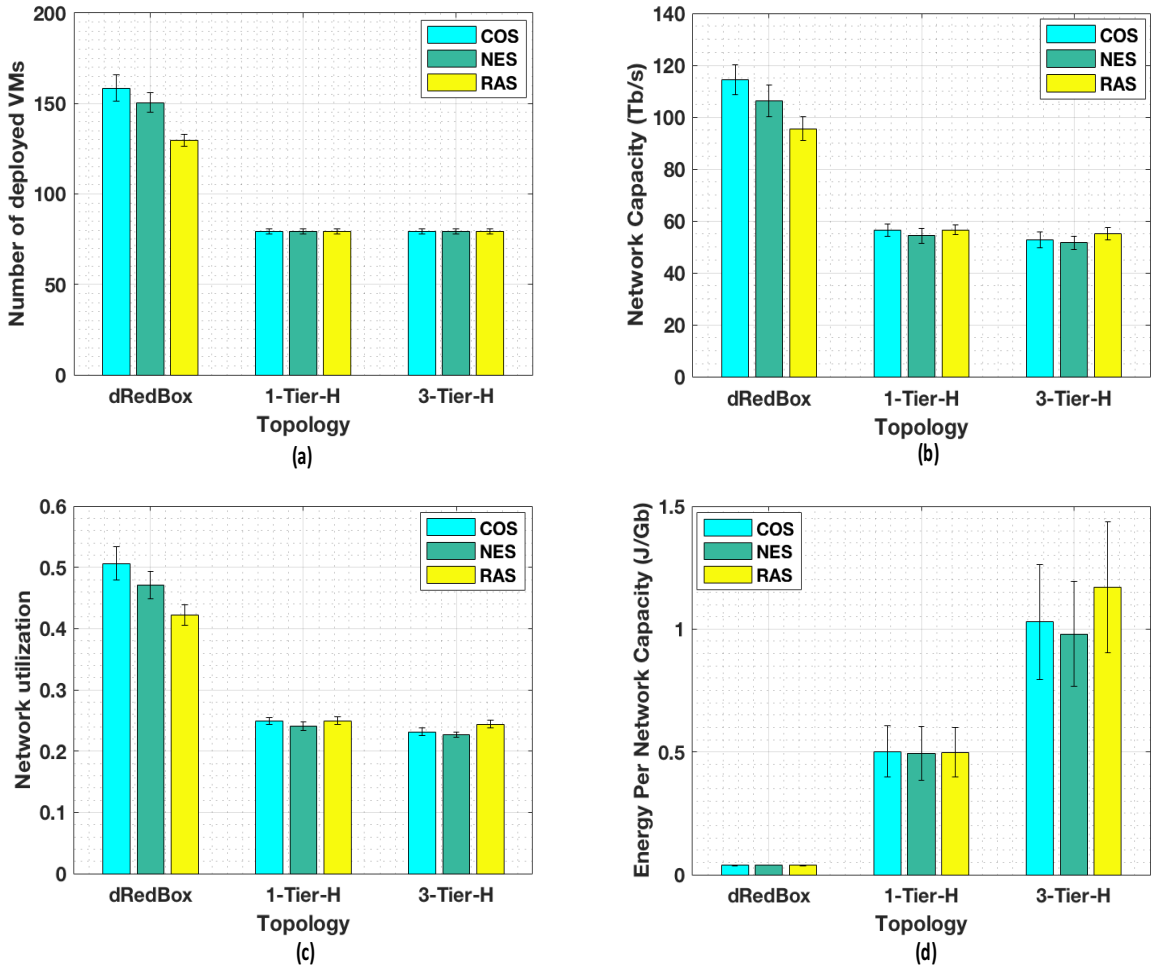

Figure 6.3: Performance indicators at 10% blocking probability for dRedBox, 1-Tier-H and 3-Tier-H architectures. (a) Number of successfully deployed VM network requests. (b) Network capacity. (c) Network utilization. (d) Energy per network capacity.

Figure 6.3(a)-(d) displays the number of successfully deployed VM network requests, network capacity, network utilization, and energy efficiency, respectively, from the 2100 holding time unit (Figure 6.2) scenario at 10% blocking probability. Numerous insights can be noted from Figure 6.3. For the dRedBox architecture, in terms of the number of successfully deployed VM network requests, D-COS has deployed approximately 5% more than D-NES and 16 % more than D-RAS. A similar trend is observed in terms of network capacity and network utilization. Furthermore, D-COS, D-NES, and D-RAS demonstrate approximately the same performance in terms of energy efficiency. This implies that D-COS demonstrates the overall best

performance, because at 10% blocking probability, D-COS has the highest number of successfully deployed VM network requests while achieving approximately the same energy efficiency with D-NES and D-RAS. For the 1-Tier-H, it is noted from Figure 6.3 that 1-T-NES has the same number of successfully deployed VM network requests with 1-T-COS and 1-T-RAS but at a slightly lower network capacity and network utilization. Also for the 3-Tier-H, it is noted from Figure 6.3 that 3-T-NES has the same number of successfully deployed VM network requests with 1-T-COS and 1-T-RAS but at a slightly lower network capacity and network utilization. Therefore, for the 1-Tier-H and 3-Tier-H architectures, the network-hop aware placement strategy uses resources more efficiently than the congestion aware and random placement strategy to achieve the same level of blocking probability. This translates into slightly better energy efficiency improvements as depicted in Figure 6.3(d). 1-T-NES demonstrates about 2% energy savings in comparison to 1-T-COS and 1% energy savings in comparison to 1-T-RAS, while 3-T-NES demonstrates about 5% energy savings in comparison to 3-T-COS and 16% in comparison to 3-T-RAS.

Comparing the performance of the dRedBox, 1-Tier-H, and 3-Tier-H architectures, in terms of the number of successfully deployed VM network requests (Figure 6.3(a)), the dRedBox architecture demonstrates a better performance than the 1-Tier-H and 3-Tier-H architectures for all EPS function placement strategies. In particular, the D-COS implemented about 100% more VM network requests than 1-T-COS, 1-T-NES, 1-T-RAS, 3-T-COS, 3-T-NES and 3-T-RAS. Furthermore, in terms of energy efficiency, dRedBox architecture performs the best, the 1-Tier-H architecture ranks second and the 3-Tier-H architecture has the worst performance for all EPS placement strategies. D-COS has energy savings of approximately 92% in comparison to 1-T-NES and 96% in comparison to 3-T-NES. This is because dRedBox is supported by a pure optical network while 1-Tier-H consist of power hungry electronic switches at the tray level, and 3-Tier-H consist of power hungry electronic switches at the tray, rack and cluster level. It can also be noted from Figure 6.3(c) that at 10% blocking probability, the 3-T-NES and 3-T-COS achieves the same number of successfully deployed VMs with the 1-T-NES but at lower network utilization. In particular, 3-T-NES demonstrates about 6% lower network utilization in comparison to 1-T-NES, while 3-T-COS demonstrates about 4% lower network utilization in comparison to 1-T-NES. The lower network utilization demonstrated by the 3-T-NES and 3-T-COS in comparison to 1-T-NES is because the 3-Tier-H architecture can perform EPS on the tray, rack and cluster networking level, while the 1-Tier-H architecture can only perform EPS on the tray level. Thus, 3-Tier-H requires less network hops to implement

EPS virtual/logical topologies than the 1-Tier-H. However, this comes at a demerit of increased power consumption and a decrease in energy efficiency (see Figure 6.3 (d)) because of multiple layers of power hungry electronic switches.

6.2.3 Proportional Bandwidth Traffic Variation

For the second simulation scenario, 2100 holding time units is used and different percentage ratios of the total generated VM network requests are assigned between LTBP and HTBP. In addition, all the points plotted in the figures presented in this section represent the points after the last VM network request, i.e., 400th request, has been received and processed in the DC network. Figure 6.4 displays the blocking probability for all architectures across different bandwidth percentage ratios. It can be observed from Figure 6.4 that D-COS, D-NES and D-RAS demonstrates lower blocking probability than the 1-Tier-H and 3-Tier-H architectures from 50:50 percentage ratio to 0:100 percentage ratio, while 1-T-COS, 1-T-NES, 1-T-RAS, 3-T-COS, 3-T-NES and 3-T-RAS demonstrates similar blocking probability across all bandwidth percentage ratios. In more detail, the D-COS demonstrates the lowest blocking probability for all bandwidth percentage ratio except at 0:100, where it is equal to D-NES and D-RAS. This is because at 0:100 bandwidth percentage ratio, all VM network requests are served by only OCS virtual/logical topologies (i.e., the EPS function placement strategies do not have any effect because no EPS virtual/logical topology is built). Additionally, it can be observed that the greater the increase in the percentage share of LBTP, the lower the blocking probability of all DC architectures and EPS placement strategies. This is because as the percentage share of LBTP increases, the greater the probability of forming EPS virtual/logical topologies, which in turns leads to conservation of more resources to accommodate future requests.

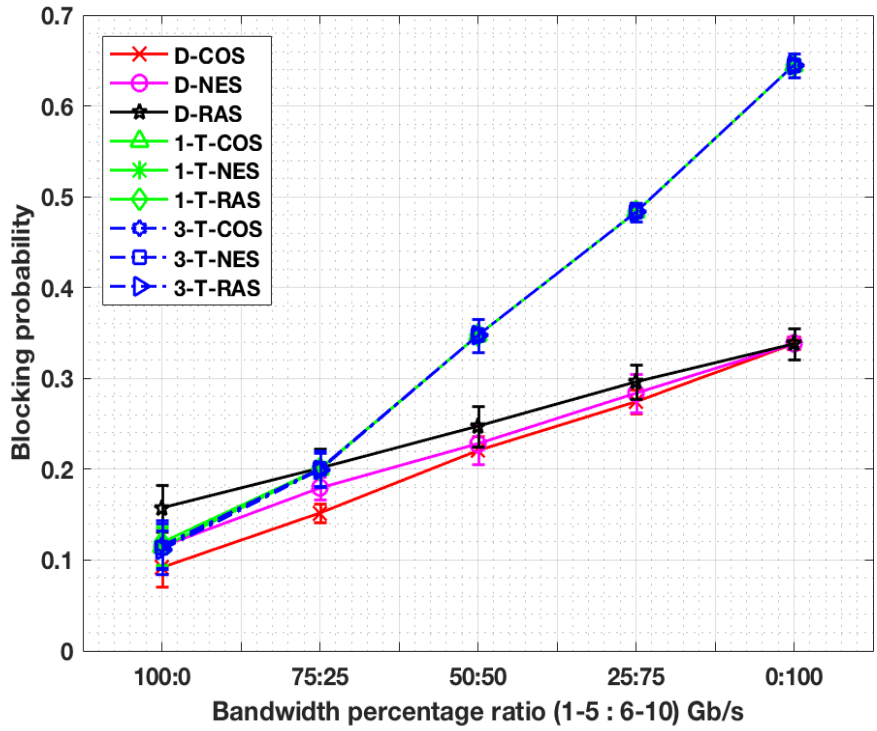

Figure 6.4: Blocking probability for proportional bandwidth traffic variation.

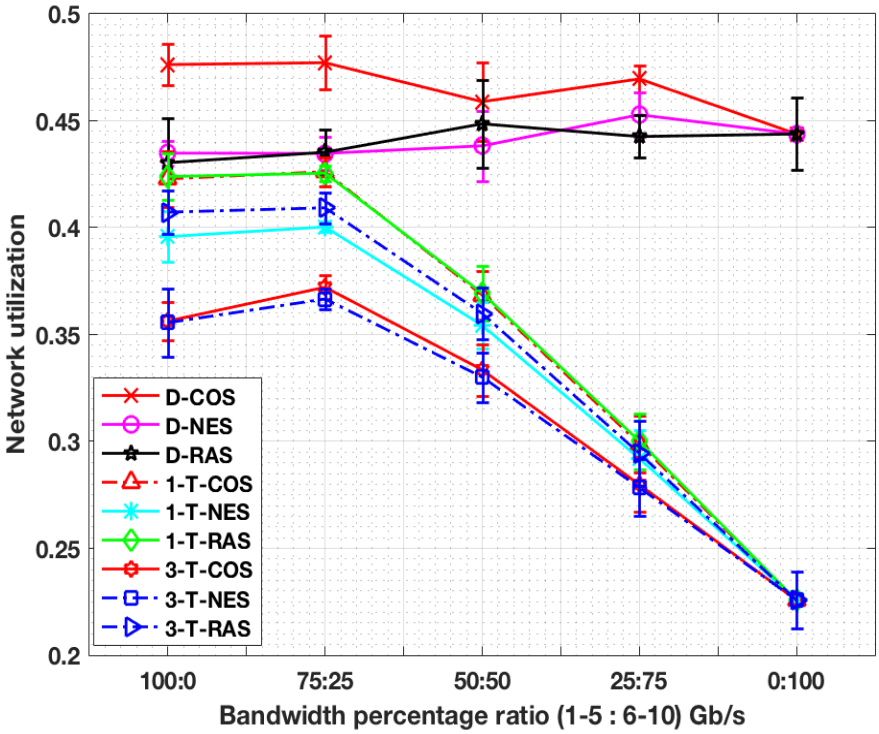

Figure 6.5: Network utilization for proportional bandwidth traffic variation

Figure 6.5 shows the network utilization across different bandwidth percentage ratios. It is noted that the network utilization of the D-COS, D-NES and D-RAS remains approximately

constant across different distributions of bandwidth percentage ratios. On the other hand, the network utilization for 1-T-COS, 1-T-NES, 1-T-RAS, 3-T-COS, 3-T-NES and 3-T-RAS decreases as the percentage share of LBTP decreases from 75% to 0%. The approximately constant utilization demonstrated by dRedBox architecture is as a result of its ability to reconfigure and deploy either packet or circuit switching services to any I/O ports to handle traffic variations. Therefore, as the bandwidth percentage ratio between LBTP to HBTP varies, the ratio between packet to circuit switch port utilization on the bricks also varies to match the traffic variations. The drop of network utilization demonstrated by the 1-Tier-H and 3-Tier-H architecture is due to the fact that the I/O ports are fixed and dedicated to either OCS or EPS. This confirms the high blocking probability demonstrated by the 1-Tier-H and 3-Tier-H architecture in Figure 6.4.

Comparing the 1-Tier-H and 3-Tier-H architectures, it can be observed from Figure 6.5 that 3-T-COS and 3-T-NES demonstrates a lower network utilization in comparison to the 1-T-NES from 100:0 to 25:75 bandwidth percentage ratio. This because the 3-Tier-H architecture can perform EPS on more networking levels (tray, rack and cluster), while the 1-Tier-H architecture can perform EPS at only the tray level. However, the benefits of more layers of EPS switches in the 3-Tier-H architecture comes at a cost of increased power consumption and a decrease in energy efficiency when compared to the 1-Tier-H architecture, which is supported by a pure optical network on the rack and cluster network layer. This evaluation confirms the trend demonstrated in Figure 6.3 and discussed in the random bandwidth variation scenario regarding the trade-off between energy and network utilization between the 1-Tier-H and 3-Tier-H architectures. Thus, the following conclusions can be made for the 1-Tier-H and 3-Tier-H architectures. First, for each of the architectures, the network-hop aware strategy uses less networking resources that than the congestion and random aware placement strategies while delivering similar levels of blocking probability. Secondly, at the same level of blocking probability, the 3-Tier-H architecture uses less networking resources than the 1-Tier-H architecture because of the advantage of having more network layers to perform EPS. However, this comes at a cost of increased power consumption which leads to a decrease in energy efficiency.

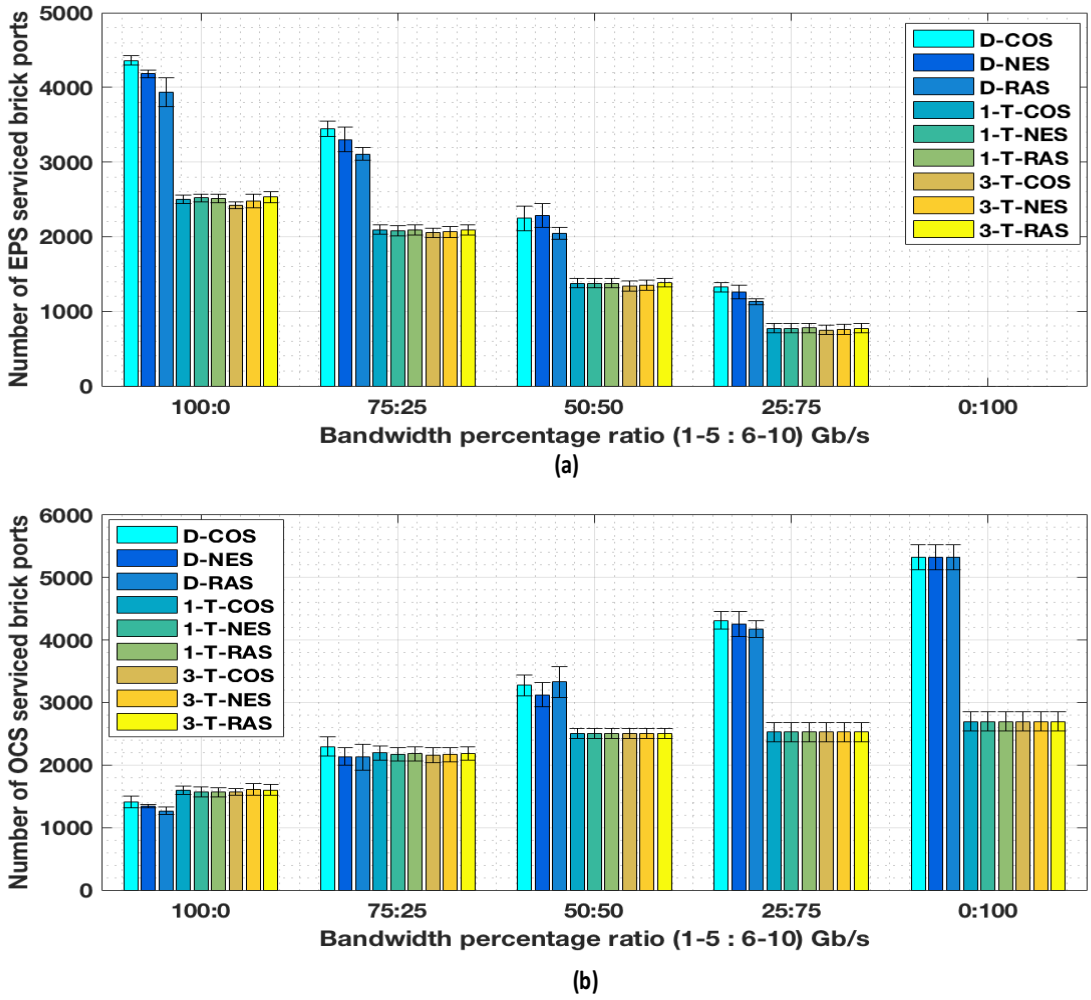

Figure 6.6: Brick port utilization for proportional bandwidth traffic variation. (a) EPS serviced ports. (b) OCS serviced ports.

Figure 6.6(a) show the number of utilized bricks ports with EPS services, while Figure 6.6 (b) shows the number of utilized bricks ports with OCS services. Figure 6.6 verifies the results of network utilization shown in Figure 6.5, and blocking probability shown in Figure 6.4. In relation to network utilization, the approximately constant network utilization displayed by the D-COS, D-NES and D-RAS can be verified by the high number of utilized brick ports across all bandwidth percentage ratios. In relation to blocking probability, the dRedBox architecture demonstrates a lower blocking probability than the 1-Tier-H and 3-Tier-H architecture from 50:50 to 0:100 bandwidth percentage ratios. This is because of the ability of the dRedBox to reconfigure any or all of the brick I/O ports to support either packet or circuit switching services.

6.2.4 Cost Analysis

Figure 6.7 displays the cost of the number of transceivers in dRedBox, 1-Tier-H and 3-Tier-H architectures across different number of racks. To evaluate the difference in cost of the different DC architectures, only the cost of transceivers is considered for two reasons. Firstly, there is no extra cost in embedding the hybrid and programmable packet/circuit switch on each brick because it is fabricated as part of the MPSoC hardware of the CPU brick where the CPUs resides or the memory brick where the memory controller resides. Instead, software is deployed on the programmable logic of the MPSoC to provide packet/circuit services [84]. Secondly, the prices of an optical switch port and electronic switch port are considered equal based on the references provided (see Table 6.1), and the total number of network ports for each of the various architectures are equal.

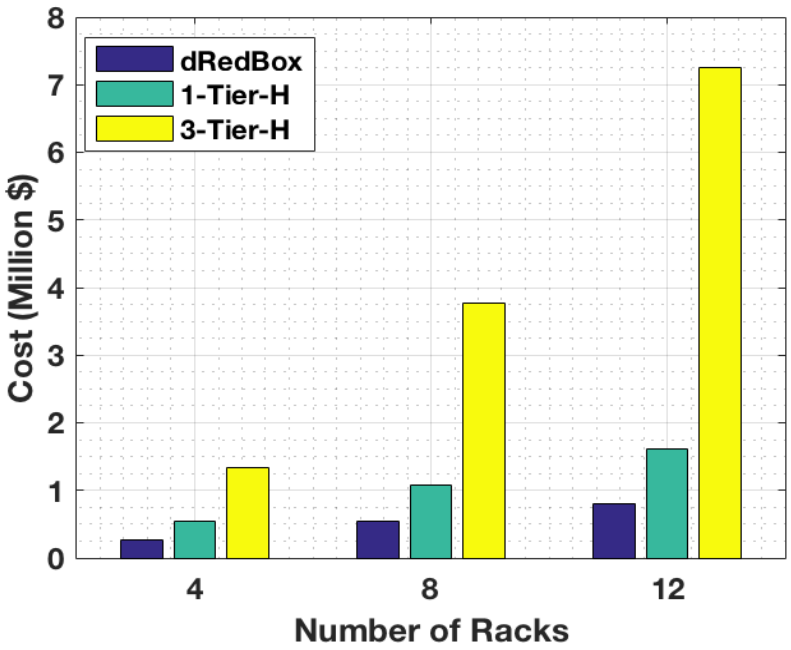

Figure 6.7: Cost of transceivers

For the dRedBox architecture, the total cost is equal to the cost of all the transceivers embedded on the bricks. For the 1-Tier-H architecture, the total cost is equal to the cost of all the transceivers embedded on the bricks and EoT electronic packet switches. For the 3-Tier-H architecture, the total cost is equal to the cost of all the transceivers embedded on the bricks, EoT electronic packet switches, ToR electronic packet switches, and ToC electronic packet switches. The results clearly show that dRedBox is the least costly architecture, followed by 1-Tier-H architecture, and 3-Tier-H architecture which is the most expensive. In more detail, the dRedBox has a 50% cost reduction in comparison to 1-Tier-H and an 80% cost reduction in

comparison to 3-Tier-H at 4 racks. The improvement in cost savings for the dRedBox architecture in comparison to the 1-Tier-H and 3-Tier-H architecture can also be observed for 8 and 12 number of racks. The reason for the significant reduction in the cost of the dRedBox is due to the additional transceivers present on the electronic switches at the tray level in 1-Tier-H, and at the tray, rack and cluster level in 3-Tier-H.

6.3 Different Arrangement of Disaggregated Resource Pools

In this section, a performance evaluation of the congestion aware and network-hop aware EPS placement strategies across different resource pools configurations in the dRedBox architecture is presented. Recalling from Chapter 6, there are three different arrangement approaches for disaggregated resource pools in DCs which include: RDC-1 (see Figure 5.1), RDC-2 (see Figure 5.2), and RDC-3 (see Figure 5.3).

6.3.1 Simulation Assumption

The simulation assumptions presented in this section are similar to the assumptions presented in the dRedBox architecture versus conventional hybrid disaggregated DC architectures simulation scenario. The only difference is the arrangement of CPU and memory resources for the RDC-2 and RDC-3.

6.3.2 Random Bandwidth Traffic Variation

Similar to the dRedBox architecture versus conventional hybrid disaggregated DC architectures simulation scenario, the VM network requests generated are randomly assigned to either LBTP or HBTP. In addition, all the points plotted in the figures presented in this section represent the points after the last VM network request, i.e., 400th request, has been received and processed in the DC network. Figure 6.8 presents the blocking probability for the congestion aware and network-hop aware EPS placement strategies across different arrangement configurations of resource pools in the dRedBox architecture and different holding times. It can be observed from Figure 6.8 that blocking probability for congestion aware placement strategy for RDC-1 (COS-RDC-1), congestion aware placement strategy for RDC-2 (COS-RDC-2) and congestion aware placement strategy for RDC-3 (COS-RDC-3) across different holding times are overlapping, suggesting similarity in terms of acceptance of VM requests. The same behaviour in blocking probability can also be observed for the network-hop aware placement strategy for RDC-1 (NES-RDC-1), network-hop aware

placement strategy for RDC-2 (NES-RDC-2) and network-hop aware placement strategy for RDC-3 (NES-RDC-3).

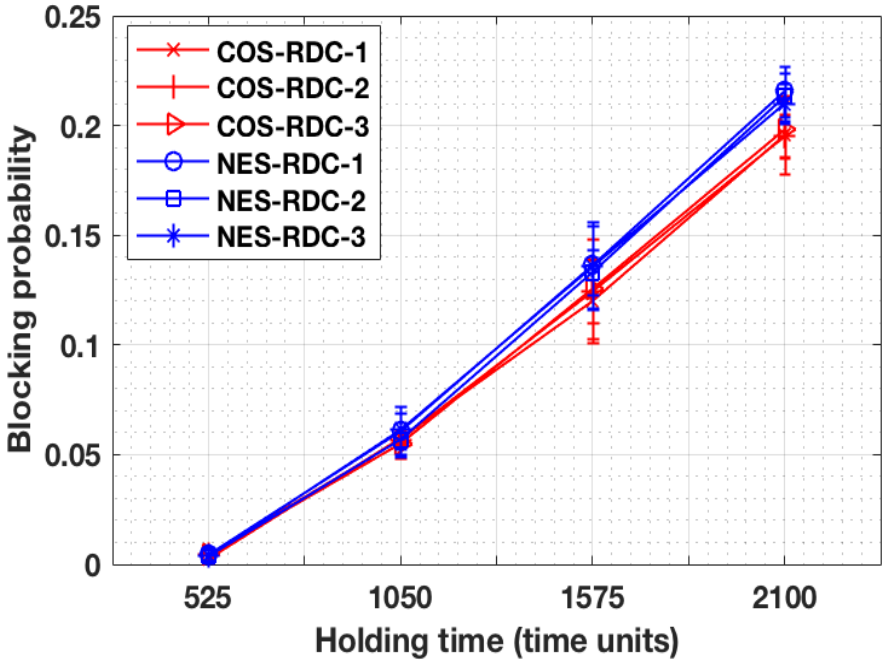

Figure 6.8: Blocking probability for different resource pool configurations

Figure 6.9(a) and Figure 6.9(b) illustrates the number of successfully deployed VM network requests, and number of utilized optical switch ports, respectively, from the 2100 holding time scenario (Figure 6.8) and at a point after the 400th request has been received in the DC. Numerous insights can be noted from Figure 6.9. It is observed that the congestion aware EPS placement strategy across the different resource pool configurations demonstrates approximately the same number of successfully deployed VM network requests. However, in terms of the number of utilized optical switch ports, the RDC-1 utilizes approximately 425 optical switch ports less than RDC-2 and 2031 optical switched ports less than RDC-3. This is because in the RDC-1, CPU to memory communication can be implemented within the same tray, rack and cluster, whereas in RDC-2, CPU to memory communication cannot be implemented within the same tray, and can only occur either within the same rack or cluster. Furthermore, in the RDC-3, CPU to memory communication cannot be implemented within the same tray or rack, and can only occur within the same cluster (i.e., RDC-3 requires the most number of network hops to implement CPU to memory communication). The network-hop aware EPS placement strategies also demonstrates approximately the same number of successfully deployed VM network requests across different resource pool configurations. However, in terms of the number of utilized optical switched ports, the RDC-1 utilizes

approximately 100 optical switch ports less than RDC-2 and 2101 optical switched ports less than RDC-3. Thus, it can be concluded that the RDC-3 for both the congestion aware and network-hop aware has the worst performance due to the fact that CPU to memory communication must always occur outside the tray and outside the rack, while the RDC-1 demonstrates the best performance.

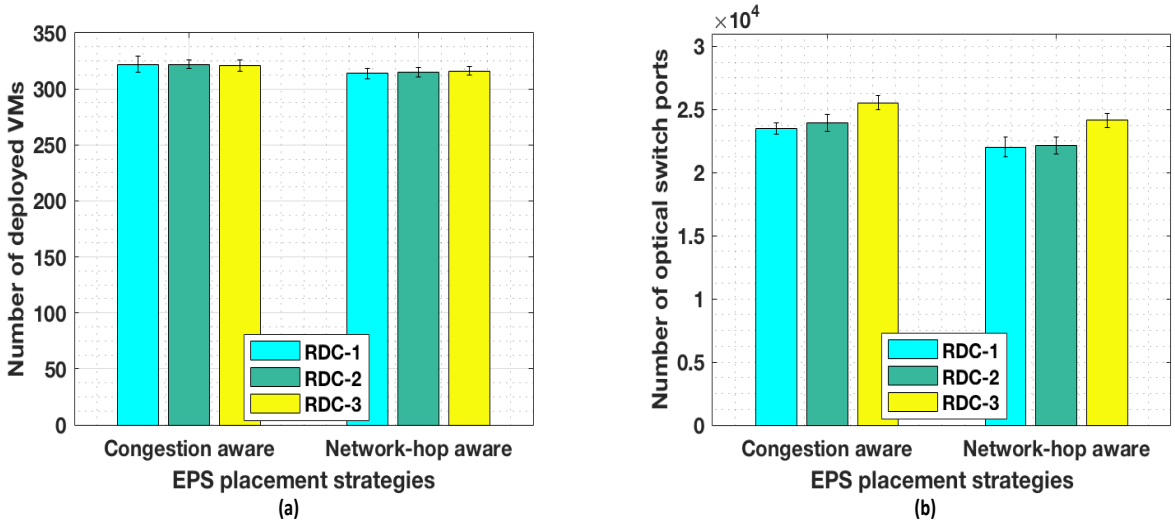

Figure 6.9: Performance indicators for RDC-1, RDC-2 and RDC-3 of dRedBox architecture. (a) Number of successfully deployed VM network requests. (b) Number of utilized optical switch ports.

6.4 Chapter Conclusion

In this chapter, the performance of the various EPS placement strategies across dRedBox, 1-Tier-H and 3-Tier-H architectures were evaluated in terms of blocking probability, number of successfully deployed VMs, network capacity, network utilization, and energy efficiency under different traffic pattern. Also, a cost analysis of the different disaggregated DC architectures was presented. Following this, a performance evaluation in terms of blocking probability, number of successfully deployed VMs and number utilized of optical switch ports for the congestion aware and network-hop aware EPS placement strategies across different resource pool configurations for dRedBox architecture was reported.

Extensive analysis and results show that for the dRedBox architecture, the congestion aware placement strategy demonstrates the best performance in terms of blocking probability. Furthermore, at 10% blocking probability, the congestion aware placement strategy delivers a

higher number of successfully deployed VM requests than the network-hop aware and random placement strategy while delivering a similar level of energy efficiency. For the 1-Tier-H and the 3-Tier-H architectures, the network-hop placement strategy performs best because it uses less network capacity and networking resources than the congestion aware and random placement strategies but demonstrates a similar performance in blocking probability. This translates to energy efficiency savings when compared to the congestion aware and random placement strategies. Furthermore, comparing the dRedBox, 1-Tier-H and 3-Tier-H architectures, the dRedBox architecture delivers substantially better performance than the 1-Tier-H and 3-Tier-H architectures. In terms of blocking probability, the dRedBox architecture demonstrated about 35% decrease in blocking probability compared to the 1-Tier-H and 3-Tier-H architecture. In terms of energy efficiency, the dRedBox architecture demonstrated 92% energy savings in comparison to the 1-Tier-H and 96% in comparison to the 3-Tier-H. In terms of cost, the dRedBox is the least expensive architecture with about 50% and 80% cost savings in comparison to the 1-Tier-H and the 3-Tier-H, respectively. Regarding the different resource pool configurations of the dRedBox, for both the congestion aware and network-hop aware strategies, the RDC-1 performs best as it utilizes less number of optical switch ports than the RDC-2 and RDC-3 while demonstrating approximately the same number of deployed VMs. The RDC-3 demonstrated the worse performance whereas the RDC-2 ranked second.

7 Conclusion and Future work

7.1 Conclusion

The main motivation of this thesis was to investigate various approaches for delivering/improving flexibility in optical metro/core and DC networks in an efficient way. This is because provisioning optical metro/core and DC networks with innovative technologies and resources to deliver/improve flexibility does not automatically translate to cost-efficiency and optimum network performance when handling dynamic and unpredictable traffic trends.

There have been numerous studies which have proposed and demonstrated approaches to deliver flexibility in optical networks. However, most of these studies did not consider flexibility as a measurable KPI. Instead, the performance of their various approaches were evaluated in terms of other KPIs such as cost and energy efficiency. Thus, there is limited understanding of how different levels of flexibility relates to other KPIs and design features. Also, the combination of server resource disaggregation, electrical/optical technologies, and network function programmability in DCs has great potential to eliminate the limitations of conventional server-centric DC architectures by providing benefits such as improved flexibility, efficient resource utilisation and energy efficiency. However, in order for this concept to materialize, resource allocation policies which ensure optimum level of networks performance must be developed. This thesis examines two different approaches to address these issues. The first approach investigates flexibility as a measurable KPI for optical switching and transmission systems. This is achieved through quantitative measurement of flexibility, and studying the relationship between flexibility, other KPIs, and design features. The second approach investigates network strategies and algorithms for the optimum placement and allocation of EPS/OCS services and network resources to build VMs on various disaggregated DC architectures. Several networking strategies and algorithms are proposed, and a comprehensive performance evaluation of the proposed algorithms across various disaggregated DC architectures and different traffic patterns was conducted.

In order to achieve the research contributions in this thesis, it was compulsory to first review existing literature and understand numerous concepts of the different aspects which are essential for delivering/improving flexibility in optical metro/core and DC networks. Chapter 2 presented a critical review of existing literature on hardware technologies, network infrastructure, software technologies such as NFV and SDN, resource allocation, and the study

of flexibility as a measurable KPI. The design principles, benefits, challenges and research gaps of the various aspects were identified and reported.

In Chapter 3, measurement models to quantify the flexibility of WSS and SSS under different design conditions were derived and proposed. The design configuration evaluated were: WSS/SSS without contention, WSS/SSS with contention, and WSS/SSS with port dimension reconfigurability. The proposed flexibility measurement models were used to measure the flexibility of the various WSS and SSS configurations, and the relationship and design trade-offs between flexibility and design features were evaluated. Results and theoretical analysis showed that design features such as port cross-connections, port dimension reconfigurability, spectrum range, equilibrium distribution between input and output port, and spectral granularity are vital design features which determine the resultant flexibility and connectivity of a WSS and SSS. The selection and combination of these design features offer different levels of flexibility and performance.

In Chapter 4, flexibility measurement models for an extensive range of BVT configurations and various optical transmission and switching subsystems (i.e., BVTs and WSS, BVTs and MCS, and BVTs and SSS) were derived and presented. The relationship and design trade-offs between flexibility, other KPIs and design features for different BVTs and optical transmission and switching subsystems configurations were evaluated and highlighted. Based on the results and design trade-off analysis presented, key design guidelines for designing BVTs and optical subsystems to achieve different levels of performance and functions were reported.

In Chapter 5, the features of a reconfigurable hybrid disaggregated architecture (i.e., dRedBox architecture) and two different conventional hybrid disaggregated DC architectures with static EPS/OCS network functions (i.e., 1-Tier-H and 3-Tier-H) were presented. Also, different arrangement configurations of disaggregated resource pools in DCs were discussed. Algorithms and network strategies for selecting EPS/OCS services, creating multi-layer custom network topologies, and allocating network resources to build VM network requests across various disaggregated DC architectures were proposed. In particular, three EPS placement strategies (i.e., congestion aware, network-hop aware and random placement strategies) with the objective of selecting and finding the most suitable brick or electronic packet switch to perform statistical multiplexing when building EPS virtual/logical topologies were reported.

Chapter 6 reported on a performance evaluation of EPS placement strategies across different disaggregated DC architectures in terms of blocking probability, number of successfully deployed VMs, network utilization, network capacity and energy efficiency. A cost evaluation on the different DC architectures was also presented. Furthermore, a performance evaluation in terms of blocking probability and number of utilized optical switch ports for different resource pool configurations of dRedBox was conducted. Results showed that the dRedBox architecture performed better than the 1-Tier-H and 3-Tier-H architectures in terms of blocking probability, energy efficiency, cost and number of successfully deployed VMs. For the dRedBox architecture, the congestion aware placement strategy performed better than the network-hop aware and random placement strategy in terms of blocking probability and number of successfully deployed VMs. For the 1-Tier-H and 3-Tier-H architectures, the network-hop aware strategy performed better than the congestion aware and random placement strategy in terms of network utilization and energy efficiency. Regarding the different resource pool configurations for dRedBox architecture, for both the congestion and network-hop aware placement strategy, the RDC-1 performs best in terms of optical switch ports utilization whereas the RDC-3 displayed the worst performance. The demonstrated benefits of the dRedBox architecture and insightful gains into the performance of the various network strategies have provided significant information and solutions to overcoming the limitations of conventional hybrid architectures and the challenges for network design of disaggregated DCs.

7.2 Future Work

The potential future research directions for the contributions reported in this thesis are described in this section.

A. Flexibility Measurement Analysis Considering Time

Time is an important factor which influences the flexibility of optical networks. For instance, the time required for an optical system to change from one connection state to another connection state impacts the level of flexibility because it determines how many connection states can achieved within a specified time frame. The relationship between flexibility and time in the context of SDN has been studied in [7]. However, the flexibility measurement methodology proposed did not consider a wide variety of factors and network characteristics. Thus, this methodology might not be applicable to other scenarios. In the context of optical networks, a methodology for measuring the flexibility of a system over a period time based on

maximum entropy has been proposed in [4]. However, the evaluation of the flexibility of the elastic optical nodes presented in [4], the add/drop banks presented in [5], and the optical components and subsystems presented in this thesis, were performed without considering transitions in time. There is limited understanding on the relationship between flexibility and time, and therefore, a potential future research direction in this research area.

B. Flexibility Measurement Analysis on the Network Layer

Another potential future direction is to extend the study of flexibility as a measurable KPI to the network layer. The research contributions presented in this thesis, and previous works [4], [5] on flexibility did not consider optical network scenarios. The suitability of using entropy as a methodology for measuring flexibility on the network level should be examined. This is because a wide variety of factors such as hardware devices, optical subsystem, network infrastructure, network topologies, software technologies, and traffic distribution influences the resultant flexibility and performance of a network. Thus, considering these factors when developing flexibility measurement models might introduce high levels of complexity.

C. Impact of Latency Constraint on the Performance of EPS Placement Strategies and dRedBox Architecture

Latency is an important design requirement for the realization of disaggregated DCs. It is critical that the latency demonstrated in dRedBox architecture when deploying VMs meets industry standards and avoids performance degradation. The network strategies and algorithms proposed and presented in this thesis did not consider the latency introduced when either EPS or OCS services are selected to build VMs. A potential future research direction will involve investigating the impact of network latency on the performance of the dRedBox and other disaggregated DC architectures. In particular, the algorithms presented in this thesis can be extended to take into account the latency introduced from both OCS and EPS technologies while considering standard latency requirements for CPU to memory communication. The total latency of a roundtrip of packet flows from source (CPU bricks) to destination (memory bricks) through either OCS or EPS technologies needs to be determined.

The latency contributed by optical circuit switching can be obtained from the data sheet of commercially available optical switches. In the case of the latency introduced by the programmable packet switch in the dRedBox, there are two possible approaches to calculate latency. The first approach is creating a network simulator for the dRedBox architecture on the packet layer with software package such as OMnet ++ [122]. This simulator can be combined

with the existing simulator in MATLAB so that both the circuit and packet layer can be monitored. Thus, the latency introduced from the flows of packets between CPU and memory through the programmable packet switch can be calculated. The second approach is the measurement of latency through an experimental setup consisting of a programmable packet switch (an example of this demonstration can be seen in [20]), and using samples of the results to develop a mathematical model which can accurately calculate latency in relation to other factors such as bandwidth and the different number of CPU and memory nodes interconnected.

D. Joint Optimisation of IT and Network Resource Allocation

Both IT and network resource allocations are required for the deployment of VMs in DCs. In this thesis, it was assumed that IT resources (i.e., CPU and memory) requirements for VM requests have been already been allocated. A complete and deeper understanding on the performance and effectiveness of the proposed algorithms and dRedBox architecture can be achieved by the combined optimum selection and allocation of IT and network resources. Either new IT resource algorithms or existing IT resource algorithms in literature can be incorporated into the MATLAB simulator built in thesis. Another potential area for future work involves the joint optimization of allocation of spectrum resources, IT resources and packet/circuit network services to build VMs across various disaggregated DCs interconnected via EONs. The dRedBox architecture can be used as a case study.

8 References

[1] Cisco, “Cisco Visual Networking Index: Forecast and Methodology, 2016-2021.”

[2] Cisco, “Cisco Global Cloud Index: Forecast and Methodology, 2016–2021.”

[3] R. P. Parker and A. Wirth, “Manufacturing flexibility: Measures and relationships,” *European Journal of Operational Research*, vol. 118, no. 3, pp. 429–449, 1999.

[4] N. Amaya, G. Zervas, and D. Simeonidou, “Introducing Node Architecture Flexibility for Elastic Optical Networks,” *Journal of Optical Communications and Networking*, vol. 5, no. 6, pp. 593–608, 2013.

[5] M. Garrich *et al.*, “Experimental Demonstration of Function Programmable Add/Drop Architecture for ROADMs [Invited],” *Journal of Optical Communications and Networking*, vol. 7, no. 2, pp. A335–A343, 2015.

[6] W. Kellerer, A. Basta, and A. Blenk, “Using a Flexibility Measure for Network Design Space Analysis of SDN and NFV,” in *IEEE Conference on Computer Communications Workshops*, 2016.

[7] W. Kellerer *et al.*, “How to Measure Network Flexibility? A Proposal for Evaluating Softwarized Networks,” *IEEE Communications Magazine*, vol. 56, no. 10, pp. 186–192, 2018.

[8] S. Gringeri, B. Basch, V. Shukla, R. Egorov, and T. Xia, “Flexible architectures for optical transport nodes and networks,” *IEEE Communications Magazine*, vol. 48, no. 7, pp. 40–50, 2010.

[9] International Telecommunication Union, “G.694.1, Spectral grids for WDM applications: DWDM frequency grid,” 2012.

[10] O. Gerstel, M. Jinno, A. Lord, and S. J. B. Yoo, “Elastic optical networking: A new dawn for the optical layer?,” *IEEE Communications Magazine*, vol. 50, no. 2, pp. S12–S20, 2012.

[11] A. Napoli *et al.*, “Next generation elastic optical networks: The vision of the European research project IDEALIST,” *IEEE Communications Magazine*, vol. 53, no. 2, pp. 152–162, 2015.

[12] S. Han, N. Egi, A. Panda, S. Ratnasamy, G. Shi, and S. Shenker, “Network support for resource disaggregation in next-generation datacenters,” in *HotNets-XII Proceedings of the Twelfth ACM Workshop on Hot Topics in Networks*, 2013, pp. 1–7.

[13] K. Katrinis *et al.*, “Rack-scale disaggregated cloud data centers: The dReDBox project vision,” in *Design, Automation & Test in Europe Conference & Exhibition (DATE)*,

2016, pp. 690–695.

[14] N. Farrington *et al.*, “Helios: a hybrid electrical/optical switch architecture for modular data centers,” *ACM SIGCOMM Computer Communication Review*, vol. 41, no. 4, pp. 339–350, 2010.

[15] G. Wang *et al.*, “c-Through : Part-time Optics in Data Centers,” *ACM SIGCOMM Computer Communication Review*, vol. 40, no. 4, pp. 327–338, 2010.

[16] K. Christodoulopoulos, D. Lugones, K. Katrinis, M. Ruffini, and D. O’Mahony, “Performance evaluation of a hybrid optical/electrical interconnect,” *Journal of Optical Communications and Networking*, vol. 7, no. 3, pp. 193–204, 2015.

[17] A. Pagès, M. P. Sanchís, S. Peng, J. Perelló, D. Simeonidou, and S. Spadaro, “Optimal Virtual Slice Composition Toward Multi-Tenancy Over Hybrid OCS / OPS Data Center Networks,” *Journal of Optical Communications and Networking*, vol. 7, no. 10, pp. 974–986, 2015.

[18] S. Peng *et al.*, “Multi-Tenant Software-Defined Hybrid Optical Switched Data Centre,” *Journal of Lightwave Technology*, vol. 33, no. 15, pp. 3224–3233, 2015.

[19] M. Imran, M. Collier, P. Landais, and K. Katrinis, “Performance evaluation of hybrid optical switch architecture for data center networks,” *Optical Switching and Networking*, vol. 21, pp. 1–15, 2016.

[20] Q. Chen, V. Mishra, and G. Zervas, “Reconfigurable computing for network function virtualization: A protocol independent switch,” in *International Conference on Reconfigurable Computing and FPGAs, ReConFig*, 2016.

[21] A. Pagès, R. Serrano, J. Perelló, and S. Spadaro, “On the benefits of resource disaggregation for virtual data centre provisioning in optical data centres,” *Computer Communications*, vol. 107, pp. 60–74, 2017.

[22] H. M. Mohammad Ali, T. E. H. El-Gorashi, A. Q. Lawey, and J. M. H. Elmirghani, “Future Energy Efficient Data Centers With Disaggregated Servers,” *Journal of Lightwave Technology*, vol. 35, no. 24, pp. 5361–5380, 2017.

[23] G. Zervas, H. Yuan, A. Saljoghei, Q. Chen, and V. Mishra, “Optically Disaggregated Data Centers With Minimal Remote Memory Latency: Technologies, Architectures, and Resource Allocation [Invited],” *Journal of Optical Communications and Networking*, vol. 10, no. 2, pp. 270–285, 2018.

[24] A. D. Papaioannou, R. Nejabati, and D. Simeonidou, “The benefits of a disaggregated data centre: A resource allocation approach,” *IEEE Global Communications Conference, GLOBECOM*, pp. 1–7, 2016.

[25] I. Tomkos, S. Azodolmolky, J. Sole-Pareta, D. Careglio, and E. Palkopoulou, "A tutorial on the flexible optical networking paradigm: State of the art, trends, and research challenges," *Proceedings of the IEEE*, vol. 102, no. 9, pp. 1317–1337, 2014.

[26] B. Collings, "Physical layer components, architectures and trends for agile photonic layer mesh networking," in *European Conference on Optical Communication*, 2009.

[27] JDSU, "A Performance Comparison of WSS Switch Engine Technologies," 2009.

[28] G. Baxter *et al.*, "Highly programmable wavelength selective switch based on liquid crystal on silicon switching elements," in *Optical Fiber Communication Conference and the National Fiber Optic Engineers Conference*, 2006.

[29] N. K. Fontaine, R. Ryf, and D. T. Neilson, "N×M Wavelength Selective Crossconnect with Flexible Passbands," in *Optical Fiber Communication Conference and the National Fiber Optic Engineers Conference*, 2012.

[30] F. Xiao and K. Alameh, "Opto-VLSI-based N×M Wavelength Selective Switch," *Optics express*, vol. 21, no. 15, pp. 18160–18169, 2013.

[31] D. M. Marom *et al.*, "Wavelength-selective 1 × K switches using free-space optics and MEMS micromirrors: Theory, design, and implementation," *Journal of Lightwave Technology*, vol. 23, no. 4, pp. 1620–1630, 2005.

[32] S. Yuan, N. Madamopoulos, R. Helkey, V. Kaman, J. Klingshirn, and J. Bowers, "Fully integrated NxN MEMS wavelength selective switch with 100% colorless add-drop ports," in *Optical Fiber Communication and National Fiber Optic Engineers Conference*, 2008.

[33] S. Li, Z. Wan, J. Xu, S. Zhong, and Y. Wu, "Wavelength-selective switch based on a polarization-independent transmission grating and a high fill-factor micromirror array," *IEEE Photonics Technology Letters*, vol. 23, no. 17, pp. 1249–1251, 2011.

[34] T. A. Strasser and J. L. Wagener, "Wavelength-selective switches for ROADM applications," *IEEE Journal on Selected Topics in Quantum Electronics*, vol. 16, no. 5, pp. 1150–1157, 2010.

[35] T. Zami and B. Lavigne, "Benefit of pure NxM WSS for optical multiflow application," in *Optical Fiber Communication Conference and National Fiber Optic Engineers Conference*, 2013.

[36] M. Jinno, H. Takara, B. Kozicki, Y. Tsukishima, Y. Sone, and S. Matsuoka, "Spectrum-efficient and scalable elastic optical path network: Architecture, benefits, and enabling technologies," *IEEE Communications Magazine*, vol. 47, no. 11, pp. 66–73, 2009.

[37] E. Hugues-Salas *et al.*, "Next generation optical nodes: The vision of the European

research project IDEALIST," *IEEE Communications Magazine*, vol. 53, no. 2, pp. 172–181, 2015.

[38] Finisar, "WaveShaper® 16000A Reconfigurable Optical Processor." [Online]. Available: https://www.finisar.com/optical-instrumentation/waveshaper-16000a.

[39] Lumentum, "TrueFlex Twin 1x35 Wavelength Selective Switch." [Online]. Available: https://www.lumentum.com/en/products/trueflex-twin-1x35-wavelength-selective-switch.

[40] T. Watanabe *et al.*, "Compact PLC-based Transponder Aggregator for Colorless and Directionless ROADM," in *Optical Fiber Communication Conference and National Fiber Optic Engineers Conference*, 2011.

[41] T. Watanabe, K. Suzuki, and T. Takahashi, "Silica-based PLC Transponder Aggregators for Colorless ," in *Optical Fiber Communication Conference and National Fiber Optic Engineers Conference*, 2012.

[42] R. Schmogrow *et al.*, "Real-time software-defined multiformat transmitter generating 64QAM at 28 GBd," *IEEE Photonics Technology Letters*, vol. 22, no. 21, pp. 1601–1603, 2010.

[43] W. Freude *et al.*, "Software-defined optical transmission," in *International Conference on Transparent Optical Networks*, 2011.

[44] J. Hilt, M. Nölle, L. Molle, M. Seimetz, and R. Freund, "32 Gbaud real-time Fpga-based multi-format transmitter for generation of higher-order modulation formats," in *International Conference on Optical Internet*, 2010.

[45] W. Shieh, H. Bao, and Y. Tang, "Coherent optical OFDM: theory and design," *Optics Express*, vol. 16, no. 2, pp. 841–859, 2008.

[46] Y. Zhang, M. O'Sullivan, and R. Hui, "Digital subcarrier multiplexing for flexible spectral allocation in optical transport network," *Optics Express*, vol. 19, no. 22, pp. 21880–21889, 2011.

[47] A. Sano *et al.*, "No-Guard-Interval Coherent Optical OFDM for 100-Gb/s/ch Long-Haul Transmission Systems," *Journal of Lightwave Technology*, vol. 27, no. 16, pp. 3705–3713, 2009.

[48] D. J. Geisler *et al.*, "Bandwidth scalable, coherent transmitter based on the parallel synthesis of multiple spectral slices using optical arbitrary waveform generation," *Optics Express*, vol. 19, no. 9, pp. 8242–8253, 2011.

[49] G. Bosco, V. Curri, A. Carena, P. Poggiolini, and F. Forghieri, "On the performance of nyquist-WDM terabit superchannels based on PM-BPSK, PM-QPSK, PM-8QAM or

PM-16QAM subcarriers," *Journal of Lightwave Technology*, vol. 29, no. 1, pp. 53–61, 2011.

[50] S. L. Woodward and M. D. Feuer, "Benefits and Requirements of Flexible-Grid ROADMs and Networks [Invited]," *Journal of Optical Communications and Networking*, vol. 5, no. 10, pp. A19–A27, 2013.

[51] M. Jinno, H. Takara, Y. Sone, K. Yonenaga, and A. Hirano, "Multiflow optical transponder for efficient multilayer optical networking," *IEEE Communications Magazine*, vol. 50, no. 5, pp. 56–65, 2012.

[52] N. Sambo *et al.*, "Next generation sliceable bandwidth variable transponders," *IEEE Communications Magazine*, vol. 53, no. 2, pp. 163–171, 2015.

[53] N. Sambo *et al.*, "Sliceable transponder architecture including multiwavelength source," *Journal of Optical Communications and Networking*, vol. 6, no. 7, pp. 590–600, 2014.

[54] S. Yan *et al.*, "Real-time ethernet to software-defined sliceable superchannel transponder," *Journal of Lightwave Technology*, vol. 33, no. 8, pp. 1571–1577, 2015.

[55] V. López *et al.*, "Finding the Target Cost for Sliceable Bandwidth Variable Transponders," *Journal of Optical Communications and Networking*, vol. 6, no. 5, pp. 476–485, 2014.

[56] Y. Li, L. Gao, G. Shen, and L. Peng, "Impact of ROADM Colorless, Directionless, and Contentionless (CDC) Features on Optical Network Performance [Invited]," *Journal of Optical Communications and Networking*, vol. 4, no. 11, pp. B58–B67, 2012.

[57] R. Jensen, A. Lord, and N. Parsons, "Colourless, directionless, contentionless ROADM architecture using low-loss optical matrix switches," in *European Conference on Optical Communication*, 2010.

[58] R. A. Jensen, "Optical switch architectures for emerging Colorless/Directionless/Contentionless ROADM networks," *Optical Fiber Communication Conference and National Fiber Optic Engineers Conference*, 2011.

[59] Y. Sakamaki *et al.*, "Experimental demonstration of colourless, directionless, contentionless ROADM using 1×43 WSS and PLC-based transponder aggregator for 127-Gbit/s DP-QPSK system," in *European Conference and Exhibition on Optical Communication*, 2011.

[60] P. N. Ji and Y. Aono, "Colorless and directionless multi-degree reconfigurable optical add/drop multiplexers," in *Wireless and Optical Communications Conference*, 2010.

[61] E. C. Magalhães *et al.*, "Node architectures for next generation ROADMs: A comparative study among emergent optical solutions," *Journal of Microwaves,*

Optoelectronics and Electromagnetic Applications, vol. 12, no. S.I 2, pp. 156–166, 2013.

[62] N. Amaya, G. S. Zervas, and D. Simeonidou, "Architecture on demand for transparent optical networks," *International Conference on Transparent Optical Networks*, 2011.

[63] M. Garrich, N. Amaya, G. S. Zervas, P. Giaccone, and D. Simeonidou, "Architecture on demand: Synthesis and scalability," in *International Conference on Optical Networking Design and Modelling*, 2012.

[64] M. Garrich *et al.*, "Architecture on Demand Design for High-Capacity Optical SDM/TDM/FDM Switching," *Journal of Optical Communications and Networking*, vol. 7, no. 1, pp. 21–35, 2015.

[65] Marija Furdek *et al.*, "Efficient Optical Amplification in Self-Healing Synthetic ROADMs," in *International Conference on Optical Network Design and Modeling*, 2014.

[66] A. Muhammad, G. Zervas, N. Amaya, D. Simeonidou, and R. Forchheimer, "Introducing flexible and synthetic optical networking: Planning and operation based on network function programmable ROADMs," *Journal of Optical Communications and Networking*, vol. 6, no. 7. pp. 635–648, 2014.

[67] N. Amaya *et al.*, "First Demonstration of Software Defined Networking (SDN) over Space Division Multiplexing (SDM) Optical Networks," in *European Conference on Optical Communication*, 2013.

[68] M. Dzanko, M. Furdek, N. A. Gonzalez, G. Zervas, B. Mikac, and D. Simeonidou, "Experimental demonstration and benefits of self-healing hard-wired and synthetic ROADMs," in *Optical Fiber Communication Conference*, 2014.

[69] O. Rival and A. Morea, "Cost-efficiency of mixed 10-40-100Gb/s networks and elastic optical networks," in *Optical Fiber Communication Conference and National Fiber Optic Engineers Conference*, 2011.

[70] B. Kozicki, H. Takara, Y. Sone, A. Watanabe, and M. Jinno, "Distance-adaptive spectrum allocation in elastic optical path network (SLICE) with bit per symbol adjustment," in *Optical Fiber Communication Conference and National Fiber Optic Engineers Conference*, 2010.

[71] A. Morea, A. F. Chong, and O. Rival, "Impact of transparent network constraints on capacity gain of elastic channel spacing," in *Optical Fiber Communication Conference and the National Fiber Optic Engineers Conference*, 2011.

[72] J. L. Vizcaíno, Y. Ye, and I. T. Monroy, "Energy efficiency in elastic-bandwidth optical

networks," in *International Conference on the Network of the Future*, 2011.

[73] E. Palkopoulou *et al.*, "Quantifying Spectrum, Cost, and Energy Efficiency in Fixed-Grid and Flex-Grid Networks [Invited]," *Journal of Optical Communications and Networking*, vol. 4, no. 11, pp. B42–B51, 2012.

[74] A. E. Ozdaglar and D. P. Bertsekas, "Routing and wavelength assignment in optical networks," *IEEE\slash ACM Transactions on Networking*, vol. 11, no. 2, pp. 259–272, 2003.

[75] B. C. Chatterjee, N. Sarma, and E. Oki, "Routing and Spectrum Allocation in Elastic Optical Networks: A Tutorial," *IEEE Communications Surveys and Tutorials*, vol. 17, no. 3, pp. 1776–1800, 2015.

[76] L. Velasco, A. Castro, M. Ruiz, and G. Junyent, "Solving routing and spectrum allocation related optimization problems: From off-line to in-operation flexgrid network planning," *Journal of Lightwave Technology*, vol. 32, no. 16, pp. 2780–2795, 2014.

[77] X. Wang, K. Kuang, S. Wang, S. Xu, H. Liu, and G. N. Liu, "Dynamic routing and spectrum allocation in elastic optical networks with mized line rates," *Journal of Optical Communications and Networking*, vol. 6, no. 12, pp. 1115–1126, 2014.

[78] K. Christodoulopoulos, I. Tomkos, and E. a Varvarigos, "Elastic Bandwidth Allocation in Flexible OFDM- based Optical Networks," *Journal of Lightwave Technology*, vol. 29, no. 9, pp. 1354–1366, 2011.

[79] Y. Yin, M. Zhang, Z. Zhu, and S. J. B. Yoo, "Fragmentation-Aware Routing, Modulation and Spectrum Assignment Algorithms in Elastic Optical Networks," in *Optical Fiber Communication Conference and National Fiber Optic Engineers Conference*, 2013.

[80] M. Dallaglio, A. Giorgetti, N. Sambo, L. Velasco, and P. Castoldi, "Routing, Spectrum, and Transponder Assignment in Elastic Optical Networks," *Journal of Lightwave Technology*, vol. 33, no. 22, pp. 4648–4658, 2015.

[81] Tencent and intel, "Tencent Explores Datacenter Resource-Pooling Using Intel ® Rack Scale Architecture (Intel ® RSA)," 2013.

[82] P. X. Gao *et al.*, "Network Requirements for Resource Disaggregation," *USENIX Symposium on Operating Systems Design and Implementation*, 2016.

[83] D. Syrivelis, A. Reale, K. Katrinis, I. Syrigos, M. Bielski, and D. Theodoropoulos, "A Software-defined Architecture and Prototype for Disaggregated Memory Rack Scale Systems," in *International Conference on Embedded Computer Systems: Architectures, Modeling, and Simulation*, 2017.

[84] A. Saljoghei *et al.*, "dRedDbox : Demonstrating Disaggregated Memory in an Optical Data Centre," in *Optical Fiber Communication Conference*, 2018.

[85] G. M. Saridis *et al.*, "EVROS: All-optical programmable disaggregated data centre interconnect utilizing hollow-core bandgap fibre," in *European Conference on Optical Communication, ECOC*, 2015.

[86] Y. Yan *et al.*, "All-Optical Programmable Disaggregated Data Centre Network Realized by FPGA-Based Switch and Interface Card," *Journal of Lightwave Technology*, vol. 34, no. 8, pp. 1925–1932, 2016.

[87] M. Al-Fares, A. Loukissas, and A. Vahdat, "A scalable, commodity data center network architecture," *ACM SIGCOMM Computer Communication Review*, vol. 38, no. 4, pp. 63–74, 2008.

[88] K. Chen *et al.*, "OSA: An optical switching architecture for data center networks with unprecedented flexibility," *IEEE/ACM Transactions on Networking*, vol. 22, no. 2, pp. 498–511, 2014.

[89] A. Singla, A. Singh, K. Ramachandran, L. Xu, and Y. Zhang, "Proteus: a topology malleable data center network," in *ACM SIGCOMM Workshop on Hot Topics in Networks (Hotnets-IX)*, 2010.

[90] K. Chen *et al.*, "WaveCube: A scalable, fault-tolerant, high-performance optical data center architecture," in *IEEE Conference on Computer Communications (INFOCOM)*, 2015.

[91] M. Dawson, "Petabit Optical Switch for Data Center Networks," in *Technical report, Polytechnic Institute of NYU*, 2010.

[92] M. Imran, M. Collier, P. Landais, and K. Katrinis, "HOSA: Hybrid Optical Switch Architecture for Data Center Networks," in *ACM International Conference on Computing Frontiers (CF '15)*, 2015.

[93] S. Peng *et al.*, "A novel SDN enabled hybrid optical packet/circuit switched data centre network: The LIGHTNESS approach," in *European Conference on Networks and Communications*, 2014.

[94] ETSI, "Network Functions Virtualisation,An Introduction, Benefits, Enablers, Challenges & Call for Action," 2012.

[95] R. Mijumbi, J. Serrat, J. L. Gorricho, N. Bouten, F. De Turck, and R. Boutaba, "Network function virtualization: State-of-the-art and research challenges," *IEEE Communications Surveys and Tutorials*, vol. 18, no. 1, pp. 236–262, 2016.

[96] Q. Chen, V. Mishra, N. Parsons, and G. Zervas, "Hardware Programmable Network

Function Service Chain on Optical Rack-Scale Data Centers," in *Optical Fiber Communication Conference*, 2017.

[97] J. Gil Herrera and J. F. Botero, "Resource Allocation in NFV: A Comprehensive Survey," *IEEE Transactions on Network and Service Management*, vol. 13, no. 3, pp. 518–532, 2016.

[98] M. Xia, M. Shirazipour, Y. Zhang, H. Green, and A. Takacs, "Network function placement for NFV chaining in packet/optical datacenters," *Journal of Lightwave Technology*, vol. 33, no. 8, pp. 1565–1570, 2015.

[99] V. Nikam, J. Gross, and A. Rostami, "VNF service chaining in optical data center networks," in *IEEE Conference on Network Function Virtualization and Software Defined Networks*, 2017.

[100] P. C. Lin, C. F. Wu, and P. H. Shih, "Optimal Placement of Network Security Monitoring Functions in NFV-Enabled Data Centers," in *IEEE 7th International Symposium on Cloud and Service Computing (SC2)*, 2017.

[101] C. H. Hsieh, J. W. Chang, C. Chen, and S. H. Lu, "Network-aware service function chaining placement in a data center," in *Asia-Pacific Network Operations and Management Symposium*, 2016.

[102] Z. Li and Y. Yang, "Placement of Virtual Network Functions in Hybrid Data Center Networks," in *IEEE 25th Annual Symposium on High-Performance Interconnects (HOTI)*, 2017.

[103] W. Fang, M. Zeng, X. Liu, W. Lu, and Z. Zhu, "Joint Spectrum and IT Resource Allocation for Efficient VNF Service Chaining in Inter-Datacenter Elastic Optical Networks," *IEEE Communications Letters*, vol. 20, no. 8, pp. 1539–1542, 2016.

[104] M. Zeng, W. Fang, and Z. Zhu, "Orchestrating tree-type VNF forwarding graphs in inter-DC elastic optical networks," *Journal of Lightwave Technology*, vol. 34, no. 14, pp. 3330–3341, 2016.

[105] Y. Wang, P. Lu, W. Lu, and Z. Zhu, "Cost-Efficient Virtual Network Function Graph (vNFG) Provisioning in Multidomain Elastic Optical Networks," *Journal of Lightwave Technology*, vol. 35, no. 13, pp. 2712–2723, 2017.

[106] X. Chen, Z. Zhu, R. Proietti, and S. J. Ben Yoo, "On Incentive-Driven VNF Service Chaining in Inter-Datacenter Elastic Optical Networks: A Hierarchical Game-Theoretic Mechanism," *IEEE Transactions on Network and Service Management*, 2018.

[107] O.N.F, "Software-defined networking: The new norm for networks," 2012.

[108] S. Sezer *et al.*, "Are We Ready for SDN ? Implementation Challenges for Software-

Defined Networks," *IEEE Communications Magazine*, vol. 51, no. 7, pp. 36–43, 2013.

[109] W. Xia, Y. Wen, C. H. Foh, D. Niyato, and H. Xie, "Survey on software-defined networking," *IEEE Communications Surveys & Tutorials*, vol. 17, no. 1, pp. 27–51, 2015.

[110] B. N. Astuto, M. Mendonca, X. N. Nguyen, K. Obraczka, and T. Turletti, "A Survey of Software-Defined Networking: Past, Present, and Future of Programmable Networks," *IEEE Communications Surveys & Tutorials*, vol. 16, no. 3, pp. 1617–1634, 2014.

[111] Y. Li and M. I. N. Chen, "Software-Defined Network Function Virtualization: A Survey," *IEEE Access*, vol. 3, pp. 2542–2553, 2015.

[112] D. Shi and R. L. Daniels, "A survey of manufacturing flexibility: Implications for e-business flexibility," *IBM Systems Journal*, vol. 42, no. 3, pp. 414–427, 2003.

[113] R. Beach, A. P. Muhlemann, D. H. R. Price, A. Paterson, and J. A. Sharp, "Review of manufacturing flexibility," *European Journal of Operational Research*, vol. 122, no. 1, pp. 41–57, 2000.

[114] M. Pérez-Pérez, A. M. Serrano-Bedia, M. C. López-Fernández, and G. García-Piqueres, "Research opportunities on manufacturing flexibility domain: A review and theory-based research agenda," *Journal of Manufacturing Systems*, vol. 48, pp. 9–20, 2018.

[115] E. Shuiabi, V. Thomson, and N. Bhuiyan, "Entropy as a measure of operational flexibility," *European Journal of Operational Research*, vol. 165, no. 3, pp. 696–707, 2005.

[116] Ö. E. Sönmez and V. T. Koç, "On Quantifying Manufacturing Flexibility: An Entropy Based Approach," in *Proceedings of the World Congress on Engineering*, 2015.

[117] "permn(V, N, K) - File Exchange - MATLAB Central." [Online]. Available: https://uk.mathworks.com/matlabcentral/fileexchange/7147-permn-v-n-k?s_tid=srchtitle.

[118] "Union of several arrays - File Exchange - MATLAB Central." [Online]. Available: https://uk.mathworks.com/matlabcentral/fileexchange/24834-union-of-several-arrays.

[119] "barwitherr(errors,varargin) - File Exchange - MATLAB Central." [Online]. Available: https://uk.mathworks.com/matlabcentral/fileexchange/30639-barwitherr-errors-varargin?w.mathworks.com.

[120] "K-Shortest Path- Yen's algorithm - File Exchange - MATLAB Central." [Online]. Available: https://uk.mathworks.com/matlabcentral/fileexchange/32513-k-shortest-path-yen-s-algorithm.

[121] "Polatis Series 6000 48xCC." [Online]. Available: http://www.polatis.com/switch-

modules-for-oem-all-optical-switch-module-solutions-original-equipment-manufactures.asp.

[122] “OMNeT++ Discrete Event Simulator.” [Online]. Available: https://omnetpp.org/.