



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ  
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ  
ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ  
ΤΟΜΕΑΣ ΣΥΣΤΗΜΑΤΩΝ ΜΕΤΑΔΟΣΗΣ ΠΛΗΡΟΦΟΡΙΑΣ  
ΚΑΙ ΤΕΧΝΟΛΟΓΙΑΣ ΥΛΙΚΩΝ

**Ανάπτυξη τεχνικών οπτικής δρομολόγησης και  
σχεδιασμός οπτικών υποσυστημάτων για ευφυή δίκτυα  
δεδομένων με δυναμική ανάθεση εύρους ζώνης**

**ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ**

**Κωνσταντίνος Μ. Τόκας**

Διπλωματούχος Ηλεκτρολόγος Μηχανικός & Μηχανικός Υπολογιστών ΕΜΠ

**Επιβλέπων:** Ηρακλής Αβραμόπουλος

Καθηγητής ΕΜΠ

Αθήνα, Οκτώβριος 2022



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ  
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ  
ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ  
ΤΟΜΕΑΣ ΣΥΣΤΗΜΑΤΩΝ ΜΕΤΑΔΟΣΗΣ ΠΛΗΡΟΦΟΡΙΑΣ  
ΚΑΙ ΤΕΧΝΟΛΟΓΙΑΣ ΥΛΙΚΩΝ

**Ανάπτυξη τεχνικών οπτικής δρομολόγησης και σχεδιασμός  
οπτικών υποσυστημάτων για ευφυή δίκτυα δεδομένων με  
δυναμική ανάθεση εύρους ζώνης**

**ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ**

**Κωνσταντίνος Μ. Τόκας**

Διπλωματούχος Ηλεκτρολόγος Μηχανικός & Μηχανικός Υπολογιστών ΕΜΠ

**Συμβουλευτική Επιτροπή: Ηρακλής Αβραμόπουλος (Επιβλέπων)**

**Νίκος Πλέρος**

**Νικόλαος Ουζούνουλου**

Εγκρίθηκε από την επταμελή εξεταστική επιτροπή την 24<sup>η</sup> Οκτωβρίου 2022.

.....  
Ηρακλής Αβραμόπουλος  
Καθηγητής Ε.Μ.Π

.....  
Νικόλαος Πλέρος  
Αναπληρωτής Καθηγητής ΑΠΘ

.....  
Νικόλαος Ουζούνουλου  
Ομότιμος Καθηγητής Ε.Μ.Π

.....  
Αθανάσιος Παναγόπουλος  
Καθηγητής Ε.Μ.Π

.....  
Γεώργιος Ματσόπουλος  
Καθηγητής Ε.Μ.Π

.....  
Νεκτάριος Κοζύρης  
Καθηγητής Ε.Μ.Π

.....  
Γεώργιος Κανέλλος  
Επίκουρος Καθηγητής ΕΚΠΑ

Κωνσταντίνος Μ. Τόκας

Διδάκτωρ Ηλεκτρολόγος Μηχανικός και Μηχανικός Υπολογιστών Ε.Μ.Π

Copyright © Κωνσταντίνος Μ. Τόκας 2022.

Με επιφύλαξη παντός δικαιώματος. All rights reserved.

Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας εργασίας, εξ ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα. Ερωτήματα που αφορούν τη χρήση της εργασίας για κερδοσκοπικό σκοπό πρέπει να απευθύνονται προς τον συγγραφέα.

Οι απόψεις και τα συμπεράσματα που περιέχονται σε αυτό το έγγραφο εκφράζουν τον συγγραφέα και δεν πρέπει να ερμηνευθεί ότι αντιπροσωπεύουν τις επίσημες θέσεις του Εθνικού Μετσόβιου Πολυτεχνείου.

*Στην Έλενα,*

*Στους γονείς μου*



# Περίληψη

Στα πλαίσια της παρούσας διδακτορικής διατριβής παρουσιάζονται τεχνικές δρομολόγησης και υλοποιούνται οπτικά υποσυστήματα με σκοπό την δυναμική ανάθεση εύρους ζώνης σε ευφυή δίκτυα δεδομένων. Πιο συγκεκριμένα και προς αυτή την κατεύθυνση παρουσιάζεται αναλυτικά μια καινοτόμα αρχιτεκτονική δικτύου για Κέντρα Δεδομένων (datacenters) στην οποία εφαρμόζονται οι παραπάνω τεχνικές δρομολόγησης και αξιοποιούνται οπτικά υποσυστήματα.

Το **ΚΕΦΑΛΑΙΟ 1** είναι ένα εισαγωγικό κεφάλαιο στο οποίο αρχικά παρουσιάζονται διάφορες αρχιτεκτονικές δικτύων που χρησιμοποιούνται στα κέντρα δεδομένων μαζί με τα πλεονεκτήματα και τα μειονεκτήματά τους. Επιπλέον, εισάγεται η έννοια του διαχωρισμού των διαθέσιμων πόρων ενός δικτύου και παρουσιάζεται η πρόκληση της υλοποίησής τους. Έπειτα, παρουσιάζονται οι βασικές αρχές της προτεινόμενης υβριδικής αρχιτεκτονικής πάνω στην οποία υλοποιούνται και εφαρμόζονται τεχνικές και υποσυστήματα που εξετάζονται στην παρούσα διατριβή. Η συγκεκριμένη αρχιτεκτονική χαρακτηρίζεται ως “υβριδική” καθώς συνδυάζει τις κλασικές ηλεκτρικές τεχνικές δρομολόγησης με αναβαθμισμένες οπτικές τεχνικές δρομολόγησης αξιοποιώντας τα πλεονεκτήματά τους με όσο το δυνατό πιο βέλτιστο τρόπο. Στα πλαίσια της εισαγωγής γίνεται και εκτενής επισκόπηση του Επιπέδου Δεδομένων (dataplane) και του Επιπέδου Ελέγχου (control plane) της προτεινόμενης υβριδικής αρχιτεκτονικής. Συνοπτικά, η αρχιτεκτονική του επιπέδου δεδομένων βασίζεται στη διασύνδεση πολλαπλών δικτυακών καψουλών - πιο απλά μεταγωγείς rods – μέσω ενός οπτικού δακτυλίου. Παράλληλα, πολλαπλές συστοιχίες υπολογιστών (racks) διασυνδέονται σε τοπολογία αστέρα στα rods μέσω των Top-of-Rack (ToR) μεταγωγέων. Το επίπεδο ελέγχου βασίζεται σε ένα αποκεντρωμένο λογισμικό ελέγχου όλων των εμπλεκόμενων υποσυστημάτων, που συνοψίζεται με τον όρο Software-defined networking.

Στο **ΚΕΦΑΛΑΙΟ 2** γίνεται εκτενής αναφορά στον σχεδιασμό των επιμέρους υποσυστημάτων που απαρτίζουν το επίπεδο δεδομένων της προτεινόμενης αρχιτεκτονικής, ξεκινώντας από τα υποσυστήματα που απαρτίζουν τον ToR switch. Το πρώτο οπτικό υποσύστημα που παρουσιάζεται είναι ο προσαρμοζόμενος πομπός του ToR switch, που βασίζεται στην γρήγορη λειτουργία ενός προσαρμοζόμενου λέιζερ που μπορεί να εναλλάσει το μήκος κύματος εκπομπής ανάλογα με τις οδηγίες του ενορχηστρωτή του δικτύου. Δίνονται λεπτομέρειες για την οπτική πηγή με πλήρη περιγραφή της ενσωμάτωσής της σε πλακέτα και τον στατικό χαρακτηρισμό του συντονισμού μήκους κύματος στη C-band. Ακολουθεί ο χαρακτηρισμός του λέιζερ υπό δυναμική λειτουργία και η εξωτερική διαμόρφωσή του στα 10Gb/s με παράλληλη αξιολόγηση όλου του πομπού. Κατόπιν γίνεται η ανάπτυξη και διαμόρφωση του δέκτη του μεταγωγέα ToR switch της προτεινόμενης υβριδικής αρχιτεκτονικής, ενώ στο τρίτο σκέλος του κεφαλαίου γίνεται αναφορά στην

ανάπτυξη ενός γρήγορου 4×4 οπτικού μεταγωγέα για εξολοκλήρου οπτική μεταγωγή εντός του ToR switch. Για λόγους πληρότητας παρουσιάζεται εν συντομία η ανάπτυξη των οπτικών υποσυστημάτων και τεχνικών δρομολόγησης εντός του pod switch, παρόλο που δεν αποτελούν μέρος της βασικής έρευνας της παρούσας διδακτορικής διατριβής. Μέσα σε αυτά τα υποσυστήματα αναφέρονται οι βασικές λειτουργίες και ιδιότητες που προσδίδουν στην αρχιτεκτονική ο Διακόπτης Καθορισμού intra- και inter- pod επικοινωνίας, ο Διακόπτης επιλογής μήκους κύματος (Wavelength Selective Switch) και ο Παθητικός Οπτικός Δρομολογητής (AWGR). Στην τελευταία υποενότητα του Κεφαλαίου, παρουσιάζεται η Πειραματική Αξιολόγηση των τεχνικών δρομολόγησης των οπτικών υποσυστημάτων και γίνεται χαρακτηρισμός των οπτικών διαδρομών: εντός pod (intra-pod) και εκτός pod (inter-pod).

Μετά τις υλοποιήσεις του κεφαλαίου 2, στο **ΚΕΦΑΛΑΙΟ 3** περιγράφονται οι διαδικασίες ενοποίησης των επιμέρους οπτικών υποσυστημάτων και η περαιτέρω αξιολόγησή τους. Πιο συγκεκριμένα περιγράφεται αναλυτικά η ενσωμάτωση του Προσαρμοζόμενου Πομπού του ToR switch με έναν μετατροπέα ψηφιακής σε αναλογική έξοδο (DAC) μέσω πλακέτας FPGA. Έπειτα αφιερώνεται μια ολόκληρη υποενότητα στον δυναμικό έλεγχο των ενοποιημένων Οπτικών Υποσυστημάτων, με αρχή την αξιολόγηση της διεπαφή επικοινωνίας μεταξύ FPGA και DAC, ενώ ακολουθεί ο χαρακτηρισμός προσαρμοζόμενου πομπού μετά την ενσωμάτωση FPGA και DAC. Στο επόμενο σκέλος αυτής της ενότητας παρουσιάζεται πλήρως η ενσωμάτωση όλων των συστημάτων ToR και pod switch και ακολουθεί η αξιολόγησή τους μέσω διαφορετικών τεχνικών δρομολόγησης. Ακολουθεί η ανάπτυξη και η πειραματική αξιολόγηση του οπτικού τμήματος του ToR switch με τη χρήση του 4×4 οπτικού διακόπτη PLZT. Τέλος, παρουσιάζεται η πλήρης ενοποίηση των υποσυστημάτων σε μια ενιαία οπτική δομή με στόχο το πρωτότυπο μίνι datacenter που παρουσιάζεται στο επόμενο κεφάλαιο.

Το **ΚΕΦΑΛΑΙΟ 4** είναι το τελευταίο κεφάλαιο που περιέχει αποτελέσματα πάνω στο βασικό θέμα της διδακτορικής διατριβής που έγινε κατά τη διάρκεια της επίδειξης του πρωτότυπου μίνι datacenter και πραγματοποιήθηκε στο Εργαστήριο Φωτονικών Επικοινωνιών (ΕΦΕ). Αρχικά γίνεται περιγραφή της αρχιτεκτονικής δομής αλλά και της κλίμακας του πρωτότυπου datacenter με εκτενή περιγραφή των στοιχείων που την απαρτίζουν. Ακολουθεί μια περιγραφική αφήγηση της «ημέρας» στη ζωή ενός πακέτου που “κυκλοφορεί” στο δίκτυο της προτεινόμενης αρχιτεκτονικής ώστε να γίνει κατανοητή όλη η αλληλουχία των διασυνδέσεων μεταξύ των επιμέρους υποσυστημάτων του πρωτοτύπου. Έπειτα, παρουσιάζονται τα αποτελέσματα της επίδειξης σε πραγματικό χρόνο στα διάφορα σενάρια end-to-end επικοινωνίας. Τα σενάρια αυτά χωρίζονται σε δύο σκέλη, το ένα αφορά σενάρια επικοινωνίας εντός του πρωτότυπου mini-DCN (Intra-DCN) και το άλλο στην επικοινωνία Inter-Datacenter μεταξύ Αθήνας (Ελλάδα) και Πίζας (Ιταλία).

Το **ΚΕΦΑΛΑΙΟ 5** περιέχει τη σύνοψη των αποτελεσμάτων της διατριβής καθώς και προτάσεις για περαιτέρω μελλοντική έρευνα δίνοντας έμφαση στην οπτική Μεταγωγή στα Κέντρα Δεδομένων, τις εναλλακτικές αρχιτεκτονικές δικτύων και τη μελέτη γρήγορων οπτικών μεταγωγέων και περισσότερων I/Os.

# Λέξεις κλειδιά

αρχιτεκτονικές κέντρων δεδομένων, οπτική δρομολόγηση, οπτική μεταγωγή, οπτικοί διακόπτες, οπτικοί μεταγωγείς, οπτικές επικοινωνίες, φωτονικές επικοινωνίες, φωτονική, οπτικά δίκτυα



# Abstract

In the context of this doctoral thesis, routing techniques are presented, and optical subsystems are implemented for the purpose of dynamic bandwidth assignment in intelligent data networks. More specifically, and in this direction, an innovative network architecture for Data Centers is presented in detail, in which the above routing techniques are applied and optical subsystems are utilized.

CHAPTER 1 is an introductory chapter that initially introduces various network architectures used in data centers along with their advantages and disadvantages. In addition, the concept of disaggregation of the available resources of a network is introduced and the challenge of their implementation is presented. Then, the basic principles of the proposed hybrid architecture are presented on which techniques and subsystems considered in this thesis are implemented and applied. This particular architecture is characterized as "hybrid" as it combines classic electrical routing techniques with upgraded optical routing techniques utilizing their advantages in the most optimal way possible. In the context of the introduction, an extensive overview of the Data Plane and the Control Plane of the proposed hybrid architecture is given. In summary, the data plane architecture is based on the interconnection of multiple network pods through an optical ring. At the same time, multiple computer arrays (racks) are interconnected in a star topology to the pods through Top-of-Rack (ToR) switches. The control layer is based on a decentralized software control of all involved subsystems, which is summarized by the term Software-defined networking.

In CHAPTER 2, extensive reference is made to the design of the individual subsystems that make up the data layer of the proposed architecture, starting with the subsystems that make up the ToR switch. The first optical subsystem presented is the adaptive transmitter of the ToR switch, based on the fast operation of an adaptive laser that can switch the emission wavelength according to the instructions of the network orchestrator. Details of the optical source are given with a full description of its on-board integration and static characterization of the C-band wavelength resonance. Following is the characterization of the laser under dynamic operation and its external configuration at 10Gb/s with a parallel evaluation of the entire transmitter. Then the development and configuration of the receiver of the ToR switch of the proposed hybrid architecture is done, while in the third part of the chapter reference is made to the development of a fast 4×4 optical switch for all-optical switching within the ToR switch. For the sake of completeness, the development of the optical subsystems and routing techniques within the pod switch is briefly presented, even though they are not part of the main research of this PhD thesis. Within these subsystems are mentioned the basic functions and properties that the intra- and inter-pod communication Determination Switch, the Wavelength Selective Switch and the Passive Optical Router (AWGR) give to the architecture. In the last subsection of the Chapter, the Experimental Evaluation of optical

subsystem routing techniques is presented and the optical paths are characterized: within pod (intra-pod) and outside pod (inter-pod).

After the implementations of chapter 2, in CHAPTER 3 the integration procedures of the individual optical subsystems and their further evaluation are described. More specifically, the integration of the Tunable Transmitter of the ToR switch with a digital to analog output converter (DAC) through an FPGA board is described in detail. An entire subsection is then devoted to the dynamic Control of the integrated Optical Subsystems, beginning with the evaluation of the communication interface between the FPGA and the DAC, followed by the tunable transmitter characterization after FPGA and DAC integration. In the next part of this section, the integration of all ToR and pod switch systems is fully presented, followed by their evaluation through different routing techniques. Next is the development and experimental evaluation of the optical part of the ToR switch using the 4×4 PLZT optical switch. Finally, the complete integration of the subsystems into a single optical structure is presented with the aim of the prototype mini datacenter which is presented in the next chapter.

CHAPTER 4 is the last chapter containing results on the main topic of the PhD thesis made during the demonstration of the prototype mini datacenter and carried out at the Photonics Communications Research Laboratory (PCRL). First, a description of the architectural structure and scale of the prototype datacenter is given, with a detailed description of the elements that make it up. What follows is a descriptive account of the "day" in the life of a packet that "runs" on the network of the proposed architecture to understand the entire sequence of interconnections between the individual subsystems of the prototype. Then, the results of the real-time demonstration in the various end-to-end communication scenarios are presented. These scenarios are divided into two branches, one concerns communication scenarios within the original mini-DCN (Intra-DCN) and the other in the Inter-Datacenter communication between Athens (Greece) and Pisa (Italy).

CHAPTER 5 contains the summary of the thesis results as well as suggestions for further future research emphasizing optical Switching in Data Centers, alternative network architectures and the study of fast optical switches and more I/Os.

## **Keywords**

Datacenter, Top-of-Rack, optical switching, optical switches, optical communications, photonics

# Πίνακας Συντμήσεων

<b>AoD</b>	Architecture on Demand
<b>AWG</b>	Arrayed Waveguide Gratings
<b>AWGR</b>	Arrayed Waveguide Gratings Router
<b>BER</b>	Bit Error Rate
<b>CDR</b>	Clock Data Recovery
<b>COTS</b>	Commercial-Off-The-Shelf
<b>CW</b>	Continuous Wave
<b>DAC</b>	Digital-to-Analog Converter
<b>DBR</b>	Distributed Bragg Reflector
<b>DCN</b>	Datacenter
<b>DCN</b>	Datacenter Network
<b>DRAM</b>	Dynamic Random Access Memory
<b>DRS</b>	Disaggregated Rack-Scale Server
<b>DSP</b>	Digital Signal Processing
<b>DWDM</b>	Dense Wavelength Division Multiplexing
<b>EAM</b>	Electro-Absorption Modulator
<b>EDFA</b>	Erbium-Doped Fiber Amplifier
<b>EPL</b>	Eclipse Public License
<b>FIFO</b>	first-in-first-out

<b>FPGA</b>	Field Programmable Gate Array
<b>FSM</b>	Finite State Machine
<b>GUI</b>	Graphical User Interface
<b>HTTP</b>	Hyper-Text Transfer Protocol
<b>IoT</b>	Internet-of-Things
<b>ITU</b>	International Telecommunication Union
<b>JSON</b>	Java Script Object Notation
<b>MAC</b>	Media Access Control
<b>MEMS</b>	Micro-Electro-Mechanical System
<b>MG-Y</b>	Modulated Grating Y-branch structure
<b>MMI</b>	multi-mode interference
<b>MZI</b>	Mach-Zehnder Interferrometer
<b>MZM</b>	Mach-Zehnder Modulator
<b>NBI</b>	Northbound Interface
<b>NIC</b>	Network Interface Card
<b>NRZ</b>	Non-Return-to-Zero
<b>OCP</b>	Open Compute Project
<b>OCS</b>	Optical Circuit Switching
<b>OF</b>	OpenFlow
<b>OPS</b>	Optical Packet Switching
<b>PCB</b>	Printed circuit board
<b>PCI</b>	Peripheral Component Interface

<b>PDF</b>	Probability Density Function
<b>PLZT</b>	(Pb,La)(Zr,Ti)O <sub>3</sub> material
<b>POSIX</b>	Portable Operating System Interface based on UNIX
<b>PPG</b>	Pulse-Pattern Generator
<b>PPG</b>	Pulse Pattern Generator
<b>PRBS</b>	Pseudo Random Binary Sequence
<b>RAM</b>	Random Access Memory
<b>REST API</b>	Representational State Transfer Application Programming Interface
<b>RF</b>	Radio Frequency
<b>RIFFA</b>	Reusable Integration Framework for FPGA Accelerators
<b>SBI</b>	Southbound Interface
<b>SDN</b>	Software Defined Network
<b>SMSR</b>	side mode suppression ratio
<b>SOA</b>	Semiconductor Optical Amplifier
<b>SPOF</b>	single point of failure
<b>TCP</b>	Transmission Control Protocol
<b>TDMA</b>	Time Division Multiple Access
<b>TEC</b>	temperature controller
<b>TH</b>	thermistor
<b>ToR</b>	Top-of-Rack
<b>TTL</b>	Transistor–transistor logic

<b>UDP</b>	User Data Protocol
<b>UHD</b>	Ultra-High Definition
<b>VLAN</b>	Virtual Local Area Network
<b>VOA</b>	Variable Optical Attenuator
<b>VOQ</b>	Virtual Output Queues
<b>VOQ</b>	Virtual Output Queues
<b>VR</b>	Virtual Reality
<b>WDM</b>	Wavelength Division Multiplexing
<b>WSS</b>	Wavelength Selective Switch

# Ευχαριστίες – Αντί Προλόγου

Ήταν Πρωτοχρονιά του 2014, ακόμα προπτυχιακός φοιτητής στην Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών του ΕΜΠ, όταν ήρθα για πρώτη φορά σε επαφή με τον κόσμο των φωτονικών σε μια πρωτοχρονιάτικη συζήτηση μεταξύ ηλεκτρολόγων. Αυτή ήταν η μέρα που μπήκε στο μυαλό μου η ενασχόληση με τον χώρο της φωτονικής και έπειτα ακολούθησε η πρώτη επίσκεψή μου στο Εργαστήριο Φωτονικών Επικοινωνιών (ΕΦΕ) και η συνάντησή μου με τον Καθηγητή Ηρακλή Αβραμόπουλο με σκοπό την εκπόνηση της προπτυχιακής μου διπλωματικής εργασίας που είχε να κάνει με τον χαρακτηρισμό κάποιων VCSELs για οπτικά καλώδια. Από την πρώτη επαφή μας δεν μπορούσα ούτε καν να φανταστώ τι πρόκειται να επακολουθήσει. Ακόμα και μετά από την ολοκλήρωση της εργασίας μου, ομολογώ ότι δεν είχα καταλάβει αφενός το εύρος της επιστήμης αφετέρου την σημασία που θα έπαιζαν στην μετέπειτα ζωή μου το ΕΦΕ και κυρίως με ο άνθρωπος Ηρακλής Αβραμόπουλος.

Νομίζω ότι θα μου είναι πολύ δύσκολο να εκφράσω επακριβώς τα συναισθήματα και την ευγνωμοσύνη που νιώθω για τον κ. Αβραμόπουλο. Θα το προσπαθήσω όμως! Ο άνθρωπος αυτός, λοιπόν, ήταν εκείνος που με βοήθησε να κλείσω τον πρώτο κύκλο μου στο Ίδρυμα ολοκληρώνοντας τις προπτυχιακές μου σπουδές αλλά ήταν και ο άνθρωπος που μου άνοιξε έναν δεύτερο και σημαντικότερο κύκλο με την εκπόνηση της διατριβής μου στο ΕΦΕ. Το ΕΦΕ δεν είναι απλά το εργαστήριο του Καθηγητή Αβραμόπουλου, στο οποίο φοιτούν επίδοξοι ερευνητές και επιστήμονες για να κάνουν την έρευνά τους και να χτίσουν βιογραφικό. Δεν είναι απλά ένας χώρος που με πολύ αγάπη και μεράκι στήθηκε από εκείνον πριν από αρκετά χρόνια. Δεν είναι απλά ένας μοναδικός χώρος εργασίας και συνεργασίας. Για μένα είναι πολλά παραπάνω, αφού το πέρασμά μου από εκεί, μου έδωσε εικόνες και εμπειρίες που με κάνουν να νιώθω εξαιρετικά τυχερός. Ο κ. Αβραμόπουλος έχει καταφέρει να δημιουργήσει έναν πολυεπίπεδο χώρο μέσα στον οποίο χτίζονται αφενός σημαντικές συνεργασίες αλλά αφετέρου και πολύ πιο σημαντικό για μένα, πολύτιμες ανθρώπινες σχέσεις και φιλίες. Οι εμπειρίες που ζεις όντας μέλος του εργαστηρίου σε ακολουθούν για πάντα. Από τα ταξίδια στο εξωτερικό μέχρι τις καθημερινές στιγμές στο γραφείο και τα πειράματα στο εργαστήριο, ο χώρος αυτός δίνει πνοή και όραμα. Εν ολίγοις, ο κύριος Αβραμόπουλος είναι ένας πολύ σημαντικός άνθρωπος για μένα και τον ευχαριστώ πολύ για όλα.

Στη συνέχεια ακολουθούν ευχαριστίες σε πολλούς ανθρώπους που η σειρά τους είναι μάλλον τυχαία, όπως μου βγήκε και ελπίζω να μην αφήσω κάποιον ή κάποια παραπονεμένο!

Στους πρώτους των ευχαριστιών δεν μπορεί να είναι άλλοι από τον Παρασκευά Μπακόπουλο και τον Δημήτρη Καλαβρουζιώτη. Ο Δημήτρης ήταν ο ηλεκτρολόγος εκείνος που την Πρωτοχρονιά του 2014 μου μίλησε για πρώτη φορά για την φωτονική και που με έφερε σε επαφή με το εργαστήριο του κ. Αβραμόπουλου. Φυσικά δεν μπορώ να παραλείψω να τον ευχαριστήσω για την σημαντική συμβολή του στην περαιτέρω παρουσία μου στο εργαστήριο και όχι μόνο. Όσο για τον Παρασκευά, νιώθω πολύ τυχερός που συνεργάστηκα στενά μαζί του στο ΕΦΕ και τον ευχαριστώ θερμά όχι μόνο για τότε αλλά και για τώρα που από άλλο μετερίζι με εμπιστεύεται και με στηρίζει.

Ένα μεγάλο ευχαριστώ στον Δημήτρη Αποστολόπουλο για την αρμονική μας συνεργασία και την σημαντική συμβολή του στην εκτόνωση των πιεστικών υποχρεώσεων της καθημερινότητας της δουλειάς. Συνεχίζω με τον Στέφανο Δρη, που παρόλο που δεν συνεργαστήκαμε στενά, αποτελεί μέχρι και σήμερα έναν από τους «μεγάλους» φίλους μου. Ευχαριστώ πάρα πολύ τον Γιάννη Λαζάρου για την εργαστηριακή εμπειρία που μου μεταλαμπάδευσε κυρίως στα πρώτα χρόνια της παρουσίας μου στο εργαστήριο. Φυσικά ευχαριστώ πολύ την Μαριλή Σπυροπούλου και τον Χάρη Ζερβό για την άψογη συνεργασία μας και που ήταν εκεί όποτε τους χρειάστηκα.

Ένα τεράστιο ευχαριστώ στον Γιάννη Γιαννούλη που εκτός από πολύτιμος και γενναιόδωρος συνεργάτης και εξαιρετικός επιστήμονας είναι ένας σημαντικός φίλος που πάντα θα έχει μια ξεχωριστή θέση στην καρδιά μου. Ο Γιάννης είναι ένας άνθρωπος που είναι πάντοτε πρόθυμος να απαντήσει σε κάθε απορία στη δουλειά, να σκύψει πάνω από κάθε πρόβλημα και φυσικά να θέσει προβληματισμούς και να πυροδοτήσει ποικίλες συζητήσεις από φιλοσοφία και επιστήμη μέχρι πολιτική και επικαιρότητα.

Στον Γιάννη Κανάκη και στον Νίκο Ηλιάδη, ένα δυνατό δίδυμο που ήταν και αυτοί που μου έδειξαν τα πρώτα εργαστηριακά μυστικά μέσα από την επίβλεψή τους κατά την προπτυχιακή μου εργασία. Ο Γιάννης εκτός από ΑΕΚτζής είναι ένας πολύ έμπειρος ερευνητής και φυσικά μια πολύ ευγενική και ήρεμη παρουσία, είναι ένας φίλος που θα είναι εκεί για να καταλάβει και να συνεισφέρει με την ίδια διάθεση στην παρέα από τα πιο σοβαρά μέχρι τα πιο χαλαρά ζητήματα. Φυσικά είναι από τους λίγους που εκτιμά τα “κρύα” αστεία μου. Ο Νίκος Ηλιάδης εκτός από επιμελής και οργανωτικός συνεργάτης είναι και φίλος μου, παρόλο που έχουν περάσει αρκετά χρόνια από τότε που δεν δουλεύουμε παρέα.

Συνεχίζω με το «έτερο» μου ήμισυ για πολλούς μήνες στο εργαστήριο. Τον πολύτιμο φίλο μου, Χρήστο Σπαθαράκη. Παρόλο που ο Χρήστος εδώ και πολλά χρόνια πλέει σε άλλες θάλασσες, άλλοτε ήρεμες κι άλλοτε φουρτουνιασμένες, η σχέση μας είναι πολύ δυνατή. Παρόλο που τον γνώριζα ήδη από τα μαθητικά μου χρόνια, η πραγματική μας φιλία γεννήθηκε μέσα στο ΕΦΕ. Δουλεύαμε μαζί σε όλα τα επιτεύγματα που περιγράφω στην διατριβή μου και ήταν πάντα εκεί να δίνει ώθηση ο ένας στον άλλον. Πραγματικά τον ευχαριστώ για όλα. Φυσικά από το “νεφελώδες” παρεάκι δεν μπορεί να λείπει ο Γιάννης



Πάτρωνας που παρόλο που δεν ήταν μέλος του ΕΦΕ, υπήρξε ιδιαίτερα σημαντικός πυλώνας στην δουλειά που κάναμε από κοινού μαζί με τον Χρήστο και τον Παρασκευά. Οι αμέτρητες ώρες στο εργαστήριο με τον Χρήστο και τον Γιάννη ήταν στιγμές έντονης πίεσης και κούρασης αλλά και καλής χημείας και συνεργασίας γι' αυτό δεν τις αλλάζω με τίποτα.

Ευχαριστώ πολύ επίσης δυο ακόμα φίλους μου, τον Νίκο Αργύρη και τον Γιάννη Πουλόπουλο. Ο Νίκος είναι πλέον ένας από τους καλύτερούς μου φίλους και φυσικά πολύ καλός συνάδελφος. Δεν έχω να πω κάτι για τις γνώσεις του στο επιστημονικό πεδίο και φυσικά δεν σταματά να με εκπλήσσει το εύρος και το βάθος των γενικών του γνώσεων. Σχετικά τώρα με τον Γιάννη, είναι ίσως από τους πιο αυθεντικούς ανθρώπους που έχω γνωρίσει που με το ξεχωριστό του γέλιο μπορεί να διασκεδάσει και τις πιο δύσκολες περιστάσεις. Αξέχαστες όλες οι εμπειρίες μαζί τους με από το ταξίδι μας στο Λος Άντζελες, στο Σαν Ντιεγκο και τον Τιχουάνα του Μεξικού μέχρι τις εξόδους μας στην Αθήνα και τα βουνά της Αττικής.

Τώρα, για την Κωνσταντίνα Καντά και τον Παναγιώτη Τουμάση ό,τι και να πω θα είναι λίγο. Είναι οι αδυναμίες μου και το ξέρουν, είναι δικοί μου άνθρωποι και το ξέρουν, τους αγαπώ ιδιαίτερα και το ξέρουν. Αυτό που δεν ξέρουν είναι ότι έπαιξαν καθοριστικό ρόλο στην καθημερινότητά μου και τους ευχαριστώ ιδιαίτερα για τις μοναδικές στιγμές που περάσαμε και περνάμε μαζί.

Ευχαριστώ πολύ και τον Χρήστο Τσώκο, που ο παράλληλος βηματισμός μας υπήρξε οριακά καρμικός. Την ίδια χρονιά «μπήκαμε» στο ΕΦΕ, την ίδια χρονιά «φύγαμε» και ολοκληρώνουμε τον κύκλο των διδακτορικών σπουδών με διαφορά ημερών. Ο Χρήστος είναι ένα ξεχωριστό παιδί και απολαμβάνω κάθε μας συνάντηση και συζήτηση. Ευχαριστώ επίσης τον Δημήτρη Ζαβιτσάνο γιατί η συνύπαρξή μας στο ΕΦΕ, αν και σύντομη ήταν εποικοδομητική και ουσιαστική.

Πριν κλείσω αυτό τον ιδιότυπο πρόλογο, θέλω να ευχαριστήσω τον Κωστή Χριστογιάννη για την εικαστική μας συνεργασία στην παραγωγή βίντεο για τα προτζεκτ και φυσικά τον Χρήστο Κουλουμέντα που για τις λίγες αλλά σημαντικές συζητήσεις μας. Ευχαριστώ θερμά τον Πάνο Γκρούμα, τον Λευτέρη Γουναρίδη, τον Βασίλη Κατοπόδη για της ωραίες εργαστηριακές στιγμές. Ευχαριστώ επίσης την Μαρία Μασσαούτη, τον Αδαμ Ραπάκη, τον Νίκο Λύρα και φυσικά τους νεότερους του εργαστηρίου: τον Αργύρη Ντάνο, την Ευρυδίκη Κυριαζή, τον Γιώργο Μπρέστα, τον Στάθη Ανδριανόπουλο και τον Γιώργο Συριόπουλο.

Τέλος, ευχαριστώ πάρα πολύ την αγαπημένη μου Βάσια Λαμπροπούλου που μαζί παιδεύομασταν με τα διάφορα γραφειοκρατικά ενώ πάντα ήταν πρόθυμη να φροντίσει κάθε ζήτημα που προέκυπτε και την Πόλυ Βλάση για την πολύτιμη βοήθειά της σε ότι χρειάστηκα.

Φεύγοντας από τον στενό πυρήνα του ΕΦΕ, θα ήθελα να ευχαριστήσω πολύ τους φίλους μου που καθ'όλη τη διάρκεια του διδακτορικού συνέβαλλαν στην αποκλιμάκωση της

έντασης της δουλειάς. Τον Στάμο, τον Δημήτρη, την Ελένη, την Μυρτώ, τον Στρατή, τον Γιάννη, τον Βαγγέλη, την Λαμπρινή, τη Ναταλία, τον Θωμά, την Ιωάννα, τον Νίκο, την Νικολέτα και φυσικά τον αγαπημένο μου κουμπάρο και κολλητό Ευάγγελο Μαυρομμάτη του Κωνσταντίνου.

Στο σημείο αυτό θέλω να ευχαριστήσω την οικογένειά μου για την συμβολή της σε αυτήν την προσπάθεια. Πρώτα και κύρια θέλω να ευχαριστήσω τους γονείς μου, την μητέρα μου Αμαλία που είναι ο πραγματικός στυλοβάτης μας και που είναι πάντα εκεί για να μας τα δώσει όλα και τον πατέρα μου που παρόλο που έφυγε πρόωρα από τη ζωή φρόντισε να μην μας λείπει τίποτα, εκτός από την παρουσία του στις σημαντικότερες στιγμές μας. Την πανέξυπνη αδερφή μου Χαρά και τον πολυμήχανο δίδυμο αδερφό μου Άγγελο που η δύναμη που μου δίνει η ενότητά μας, με κάνει να νιώθω άτρωτος. Φυσικά, δεν μπορώ να παραλείψω την Νεκταρία που μεγαλώνει τον πιο γλυκό και όμορφο ανιψιό του κόσμου, τον μικρό μας Μάριο! Φυσικά ευχαριστώ τη θεία μου τη Νάσα και την αδυναμία μου, την γιαγιά μου την Αντιγόνη που η δύναμή της ακόμα και σήμερα στα 98 χρόνια είναι αξιοθαύμαστη.

Κλείνοντας θέλω να ευχαριστήσω τον άνθρωπό μου, την Έλενα. Νιώθω τυχερός που τη συνάντησα και που από τότε χτίζουμε μαζί το μέλλον και τη ζωή μας. Δεν είναι απλώς η σύζυγός μου, είναι αυτή που με την δύναμή της και την κατανόησή της με συμπληρώνει και που μαζί δημιουργούμε μοναδικές στιγμές και εμπειρίες ο ένας στον άλλον.

Υ.Γ. Υπάρχει περίπτωση να έχω ξεχάσει κάποιους και γι' αυτό απολογούμαι και εύχομαι να μην μου κρατήσουν κακία γιατί δεν ήταν εκούσιο!

# Πίνακας Περιεχομένων

Περίληψη .....	1
Abstract .....	4
Πίνακας Συντμήσεων .....	6
Ευχαριστίες – Αντί Προλόγου .....	10
Πίνακας Περιεχομένων .....	14
ΚΕΦΑΛΑΙΟ 1 Εισαγωγή .....	17
1.1 Αρχιτεκτονικές Δικτύων για Κέντρα Δεδομένων .....	18
1.1.1 Τοπολογίες Fat-Tree και Clos στα Δίκτυα Κέντρων Δεδομένων .....	18
1.1.2 Η πρόκληση του διαχωρισμού των πόρων (disaggregation) .....	22
1.1.3 Αρχιτεκτονικές οπτικής μεταγωγής για Δίκτυα Κέντρων Δεδομένων .....	24
1.2 Εισαγωγή στην υβριδική αρχιτεκτονική .....	27
1.3 Η δομή και τα στοιχεία της προτεινόμενης αρχιτεκτονικής .....	30
1.3.1 Επισκόπηση του Επιπέδου Δεδομένων (Data-plane) .....	31
1.3.2 Επισκόπηση του Επιπέδου Ελέγχου (Control Plane) .....	37
ΚΕΦΑΛΑΙΟ 2 Σχεδιασμός Οπτικών Υποσυστημάτων .....	40
2.1 Προσαρμοζόμενος Πομπός στο ToR switch .....	40
2.1.1 Επισκόπηση και αρχή λειτουργίας της προσαρμοζόμενης οπτικής πηγής λείζερ (Tunable Laser) .....	41
2.1.2 Ρύθμιση λείζερ και ενσωμάτωση σε πλακέτα (PCB) .....	43
2.1.3 Στατικός χαρακτηρισμός του συντονισμού μήκους κύματος του προσαρμοζόμενου λείζερ στη C-band .....	46
2.1.4 Χαρακτηρισμός ρύθμισης μήκους κύματος του λείζερ υπό δυναμική λειτουργία .....	51

2.1.5	Εξωτερική διαμόρφωση του προσαρμοζόμενου λέιζερ στα 10Gb/s με παράλληλη αξιολόγηση του πομπού.....	58
2.2	Ανάπτυξη και διαμόρφωση του δέκτη του μεταγωγέα ToR switch της προτεινόμενης υβριδικής αρχιτεκτονικής.....	61
2.3	Ανάπτυξη γρήγορου 4×4 οπτικού μεταγωγέα για εξολοκλήρου οπτική μεταγωγή εντός του ToR.....	68
2.4	Ανάπτυξη οπτικών υποσυστημάτων και τεχνικών δρομολόγησης εντός του pod switch.....	74
2.4.1	Διακόπτης Καθορισμού intra- και inter- pod επικοινωνίας.....	75
2.4.2	Διακόπτης επιλογής μήκους κύματος (Wavelength Selective Switch).....	76
2.4.3	Παθητικός Οπτικός Δρομολογητής (AWGR).....	77
2.5	Πειραματική Αξιολόγηση των τεχνικών δρομολόγησης των οπτικών και χαρακτηρισμός των οπτικών διαδρομών.....	78
2.5.1	Πειραματική αξιολόγηση της δρομολόγησης εντός pod (intra-pod).....	78
2.5.2	Πειραματική απόδειξη της δρομολόγησης εκτός pod (inter-pod).....	80
<b>ΚΕΦΑΛΑΙΟ 3 Ενοποίηση των Οπτικών Υποσυστημάτων και Αξιολόγηση Τεχνικών Δρομολόγησης</b>		
		86
3.1	Ενσωμάτωση Προσαρμοζόμενου Πομπού με μετατροπέα ψηφιακής σε αναλογική έξοδο μέσω FPGA.....	86
3.2	Δυναμικός Έλεγχος Οπτικών Υποσυστημάτων.....	91
3.2.1	Διεπαφή επικοινωνίας μεταξύ FPGA και DAC.....	94
3.2.2	Χαρακτηρισμός προσαρμοζόμενου πομπού μετά την ενσωμάτωση FPGA και DAC	98
3.3	Ενσωμάτωση συστημάτων ToR και pod switch.....	106
3.3.1	Αξιολόγηση της ενσωμάτωσης των υποσυστημάτων μέσω διαφορετικών τεχνικών δρομολόγησης.....	106
3.3.2	Αξιολόγηση μέσω Intra-pod δρομολόγησης.....	110
3.3.3	Αξιολόγηση μέσω Inter-pod δρομολόγησης.....	112
3.3.4	Αξιολόγηση μέσω Inter-pod δρομολόγησης με έμφαση στο WSS.....	115
3.4	Ανάπτυξη και πειραματική αξιολόγηση του οπτικού τμήματος του ToR switch με τη χρήση του 4×4 οπτικού διακόπτη PLZT.....	118
3.4.1	Επικοινωνία εντός του ίδιου ToR (Intra-ToR) με εξολοκλήρου οπτικό τρόπο.	119
3.4.2	Επικοινωνία σε ToR εντός του ίδιου pod με εξολοκλήρου οπτικό τρόπο....	122
3.4.3	Επικοινωνία μεταξύ ToRs σε διαφορετικά pod με εξολοκλήρου οπτικό τρόπο	126
3.5	Ενοποίηση των υποσυστημάτων σε μια ενιαία οπτική δομή.....	132
<b>ΚΕΦΑΛΑΙΟ 4 Το πρωτότυπο μίνι datacenter στο ΕΦΕ</b>		
		138

4.1	Η αρχιτεκτονική δομή και η κλίμακα του πρωτότυπου datacenter .....	138
4.2	Μια «ημέρα» στη ζωή ενός πακέτου .....	142
4.3	Σενάρια end-to-end επικοινωνίας και αποτελέσματα σε πραγματικό χρόνο .....	144
4.3.1	Σενάρια Επικοινωνίας εντός του πρωτότυπου mini-DCN (Intra-DCN).....	144
4.3.2	Επικοινωνία Inter-Datcenter μεταξύ Αθήνας (Ελλάδα) και Πίζας (Ιταλία)	160
ΚΕΦΑΛΑΙΟ 5 Σύνοψη των αποτελεσμάτων και προτάσεις για μελλοντική έρευνα .		
5.1	Σύνοψη των αποτελεσμάτων.....	162
5.1.1	Υλοποίηση και πειραματική αξιολόγηση του γρήγορου προσαρμοζόμενου πομπού του ToR switch.....	163
5.1.2	Χαρακτηρισμός και πειραματική αξιολόγηση των γρήγορων οπτικών μεταγωγέων .....	165
5.1.3	Ενοποίηση και πειραματική αξιολόγηση οπτικών υποσυστημάτων του πρωτότυπου μίνι-DataCenter στο Εργαστήριο Φωτονικών Επικοινωνιών (ΕΦΕ) ...	166
5.2	Προτάσεις για μελλοντική έρευνα .....	171
5.2.1	Οπτική Μεταγωγή στα Κέντρα Δεδομένων και εναλλακτικές αρχιτεκτονικές δικτύων.....	171
5.2.2	Μελέτη γρήγορων οπτικών μεταγωγέων και περισσότερων I/Os. ....	172
Βιβλιογραφία .....		174
Λίστα Δημοσιεύσεων.....		182
Λίστα Σχημάτων .....		186
Λίστα Πινάκων .....		194
Extended Abstract.....		195

# ΚΕΦΑΛΑΙΟ 1

## Εισαγωγή

Στις μέρες μας, τα Κέντρα Δεδομένων (Datacenters - DC) είναι η καρδιά όλων των διαδικτυακών εφαρμογών και των υπηρεσιών του Internet-of-Things (IoT), που χειρίζονται τεράστιες ποσότητες ψηφιακών δεδομένων κάθε στιγμή. Ο συνεχής εμπλουτισμός αυτών των διαδικτυακών υπηρεσιών και εφαρμογών ανοίγει νέους ορίζοντες στην εμπειρία των χρηστών και εκτοξεύει τη ζήτηση για περισσότερο εύρος ζώνης και γρηγορότερες ταχύτητες. Όλο και περισσότερες ψηφιακές υπηρεσίες που εκτείνονται από την εικονική πραγματικότητα (Virtual Reality - VR) και τη ζωντανή ροή βίντεο υψηλής ευκρίνειας (Ultra-High Definition – UHD live streaming) έως τα cloud δίκτυα αποθήκευσης και αισθητήρων, σχηματίζοντας το IoT, ξεφυτρώνουν διαρκώς σε όλο τον κόσμο, επιβαρύνοντας με τεράστιο ψηφιακό φορτίο τους κόμβους του Διαδικτύου.

Στην εποχή που το 5G κυριαρχεί και με το 6G προ των πυλών, όλες αυτές οι δραστηριότητες χρειάζονται όλο και περισσότερο εύρος ζώνης, καθώς οι χρήστες απαιτούν άμεση και on-demand πρόσβαση σε αυτές, ανά πάσα στιγμή. Αυτό έχει σαν αποτέλεσμα τα δίκτυα κέντρων δεδομένων (Datacenter Networks - DCN) να είναι ικανά να παρέχουν διασυνδέσεις εξαιρετικά υψηλής χωρητικότητας για έναν τεράστιο αριθμό κόμβων και κεντρικών υπολογιστών. Παράλληλα θα πρέπει οι διασυνδέσεις αυτές να παρουσιάζουν αφενός χαμηλή καθυστέρηση (low latency) για την εκπλήρωση απαιτητικών χρονικά υπηρεσιών και αφετέρου υψηλής απόδοσης αξιοπιστία (reliability) που αφορά στον χρόνο διακοπής των υπηρεσιών (service interruption-time). Ως εκ τούτου, αυτές οι απαιτήσεις έχουν άμεσο αντίκτυπο στην λειτουργία των Κέντρων Δεδομένων, οδηγώντας την παγκόσμια IP κίνηση του δικτύου κέντρων δεδομένων σε μια εξαιρετικά απότομη καμπύλη ανάπτυξης που φτάνει το 25% ετησίως [1].

Δεδομένου ότι η κίνηση μέσα σε ένα datacenter είναι πολύ υψηλότερη από την εισερχόμενη ή εξερχόμενη κίνηση [1], τα δίκτυα στο εσωτερικό τους αντιμετωπίζουν σημαντικές προκλήσεις σχετικά με τη χρήση των διαθέσιμων πόρων, την επεκτασιμότητα και την ευελιξία διαχείρισης. Για να αντιμετωπιστεί με επιτυχία αυτή η αυξανόμενη ζήτηση και να αποφευχθεί κάθε πιθανή κρίση χωρητικότητας ενώ παράλληλα αυξάνεται η κατανάλωση ενέργειας, οι διαχειριστές και οι πάροχοι εξοπλισμού των Κέντρων Δεδομένων αγωνίζονται να αναβαθμίσουν τις υπάρχουσες υποδομές. Τα υπερσύγχρονα δίκτυα εσωτερικά των

datacenters (intra-DCN) βασίζονται σε ηλεκτρονικούς διακόπτες/μεταγωγείς ή δρομολογητές που διασυνδέονται σε τοπολογία fat-tree ή folded-clos που χρησιμοποιούν οπτικές ίνες, με ηλεκτρο-οπτο-ηλεκτρική μετατροπή σε κάθε κόμβο [2]. Ωστόσο, οι τοπολογίες fat-tree τείνουν να υποχρησιμοποιούν τους πόρους και ταυτόχρονα απαιτούν πληθώρα καλωδίων, ινών και διακοπών.

Σε αυτό το πλαίσιο, οι τεχνολογίες οπτικής μεταγωγής (optical switching technologies) κερδίζουν έδαφος ως μια πιθανή λύση για την αντιμετώπιση των προαναφερθέντων προκλήσεων. Η χρησιμοποίηση φωτονικών στοιχείων θα μπορούσε να προσφέρει επεκτασιμότητα (scalability) στο δίκτυο, λόγω των εγγενών ιδιοτήτων τους, όπως η ταχύτητα, η ενεργειακή απόδοση και η διαφάνεια στο επίπεδο του πρωτόκολλου και του ρυθμού μετάδοσης (bitrate). Έχουν προταθεί αρκετές δουλειές που εμφανίζουν τις οπτικές τεχνολογίες [3] ως αποτελεσματικές λύσεις στα Δίκτυα των Κέντρων Δεδομένων, μεταξύ αυτών:

- α) Η χωρική μεταγωγή (space switching) για παράδειγμα με χρήση μικροηλεκτρομηχανικών συστημάτων (Micro Electro-mechanical systems - MEMS) [4] ή με χρήση οπτικών ενισχυτών ημιαγωγών (Semiconductor Optical Amplifier – SOA) [5],
- β) Η μεταγωγή μήκους κύματος (wavelength switching), μέσω συνδυασμού προσαρμοζόμενων λέιζερ και οπτικών δρομολογητών — AWGRs [6], [7]
- γ) Ο συνδυασμός των δύο παραπάνω χρησιμοποιώντας π.χ. διακόπτες επιλογής μήκους κύματος (Wavelength Selective Switches – WSS) [8]

Μία από τις βασικές προκλήσεις που επίσης αντιμετωπίζουν τα σημερινά δίκτυα οπτικών κέντρων δεδομένων είναι ο συνδυασμός αφενός της επεκτασιμότητας (scalability) του δικτύου και αφετέρου της γρήγορης επαναδιαμόρφωσης (reconfigurability). Για το σκοπό αυτό, αρκετές εργασίες προωθούν την ενσωμάτωση της οπτικής μεταγωγής σε πλατφόρμες κεντρικού ελέγχου και προγραμματισμού του δικτύου μέσω λογισμικού, που αναφέρεται με τον αγγλικό όρο Software-Defined Networking – SDN [9], [10]. Πράγματι, οι πλατφόρμες SDN σε συνδυασμό με αλγόριθμους ρύθμισης του δικτύου παρέχουν δυναμικά χαρακτηριστικά αλλά και επεκτασιμότητα στα DCN, ενώ παράλληλα ενισχύουν τα πλεονεκτήματα των οπτικών διακοπών/μεταγωγέων.

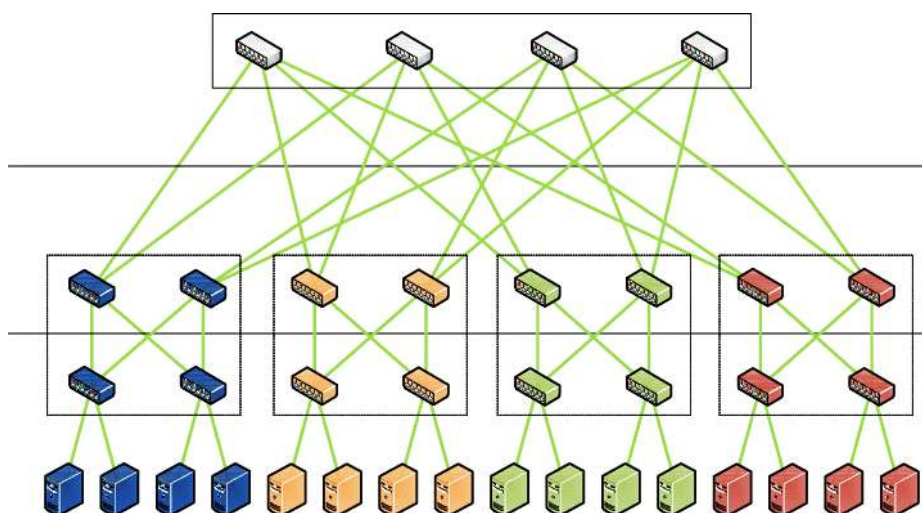
## **1.1 Αρχιτεκτονικές Δικτύων για Κέντρα Δεδομένων**

### **1.1.1 Τοπολογίες Fat-Tree και Clos στα Δίκτυα Κέντρων Δεδομένων**

Τα σημερινά Δίκτυα των Κέντρων Δεδομένων (Datacenter Networks) σχεδιάζονται βασισμένα στην fat-tree τοπολογία διασύνδεσης ή σε υπερπληθή (over-subscribed) fat-tree

τοπολογία, όπως φαίνεται στην Εικόνα 1. Πιο συγκεκριμένα ακολουθούνται δύο προσεγγίσεις στα δίκτυα τοπολογίας τύπου fat-tree:

- (i) Η παραδοσιακή προσέγγιση της fat-tree τοπολογίας που χρησιμοποιεί θύρες/διεπαφές υψηλών ρυθμών μετάδοσης με κατεύθυνσή προς τη ρίζα του δέντρου,
- (ii) Η προσέγγιση σύμφωνα με την οποία χρησιμοποιούνται εμπορικοί διακόπτες/μεταγωγείς χαμηλής ταχύτητας σε μια τοπολογία Clos (Εικόνα 2) έτσι ώστε να δημιουργούνται πολλαπλές διαδρομές χαμηλού ρυθμού μετάδοσης μεταξύ των endhost (διακομιστών) [11].

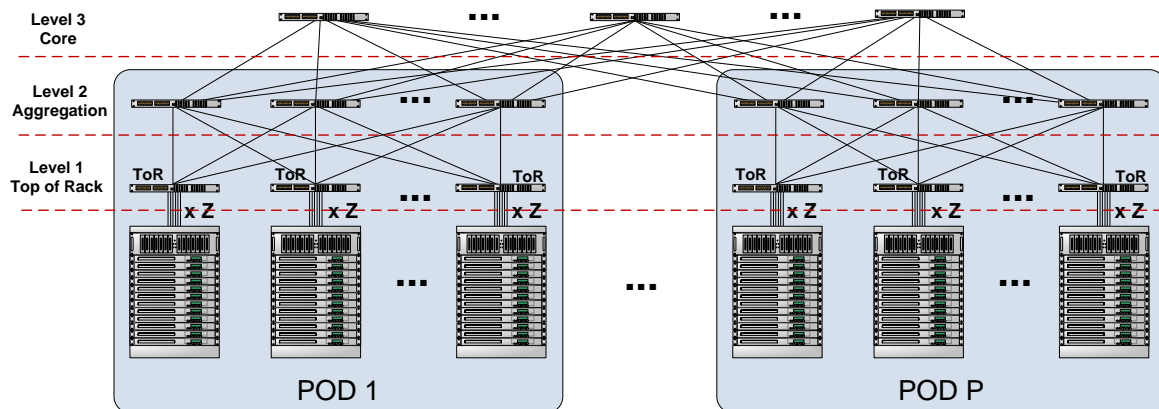


**Εικόνα 1:** Μια κλασική fat-tree τοπολογία [12]

Η πρώτη προσέγγιση στην εφαρμογή της αποδεικνύεται μη πρακτική στα δίκτυα των μεγάλων κέντρων δεδομένων, καθώς η υποστήριξη πολλών διακομιστών (endhost) απαιτεί κόμβους και διεπαφές μεγάλου εύρους ζώνης στη ρίζα του δέντρου, ενώ το κόστος αυξάνεται κατά πολύ καθώς μεταβαίνουμε από απλά εξαρτήματα χαλκού (όπως τα κλασικά καλώδια) σε πιο εξελιγμένο high-end εξοπλισμό. Επιπλέον, ένα δίκτυο που έχει δημιουργηθεί με αυτόν τον τρόπο δεν μπορεί εύκολα να επεκταθεί για να υποστηρίξει περισσότερους endhost, ενώ πολύ συχνά εμφανίζονται προβλήματα αστοχίας (single point of failure – SPOF), δηλαδή ένα και μόνο πρόβλημα είναι ικανό να δημιουργήσει διαδοχικά προβλήματα σε ολόκληρο το σύστημα. Έτσι, η πιο δημοφιλής προσέγγιση στα σημερινά δίκτυα των κέντρων δεδομένων είναι είτε η προσέγγιση Clos, η οποία περιορίζει το κόστος αφού χρησιμοποιεί πολλά αλλά φθηνά εξαρτήματα του εξοπλισμού στα υψηλότερα επίπεδα διασύνδεσης ενώ αξιοποιεί πολλαπλές διαδρομές που λύνουν προβλήματα SPOF, είτε ένας συνδυασμός από τις δύο προσεγγίσεις (τοπολογία Clos με μεταγωγείς υψηλότερου ρυθμού



μετάδοσης από αυτόν του διακομιστή). Η Εικόνα 2 δείχνει ένα δίκτυο σε τοπολογία fat-tree που έχει δημιουργηθεί από διακόπτες χαμηλής τιμής σε μια τοπολογία Clos.



**Εικόνα 2:** Fat-tree δίκτυο κατασκευασμένο από μεταγωγείς σε μια τοπολογία Clos.

Τα σημερινά κέντρα δεδομένων εξυπηρετούν δεκάδες έως εκατοντάδες χιλιάδες διακομιστές (servers/endhosts) τοποθετημένους σε rack, που το καθένα τυπικά εξυπηρετεί από 20 έως 40 servers). Ένας αριθμός από racks που εξυπηρετούν τους servers μπορεί να ομαδοποιηθεί λογικά ή φυσικά σε pods. Λογικά σημαίνει ότι τα racks ανήκουν στο ίδιο υποδίκτυο και διασυνδέονται με τους ίδιους pod switches (δηλαδή στο δεύτερο επίπεδο του δέντρου), ενώ φυσικά σημαίνει ότι τοποθετούνται στον ίδιο χώρο-κοντέινερ. Εταιρείες όπως η HP, η IBM, η Sun, κ.λπ. συσκευάζουν servers και εξοπλισμό δικτύωσης σε κλασικά κοντέινερ μεταφορών (shipping containers) που μπορούν να χρησιμοποιηθούν ως αυτόνομα μικρά κέντρα δεδομένων ή μπορούν να συνδεθούν ως αυτοτελείς μονάδες ενός μεγαλύτερου κέντρου δεδομένων. Ο αριθμός των racks ανά φυσικό pod είναι συνήθως μεταξύ 40 και 60, αλλά αυτός ο αριθμός μπορεί να είναι ακόμα μεγαλύτερος.

Κάθε rack είναι συνήθως εξοπλισμένο με έναν μεταγωγέα στην κορυφή του, που ορίζεται με τον αγγλικό όρο Top-of-Rack (ToR) switch, και διαχειρίζεται τις διεπαφές νότιας κατεύθυνσης (southbound interfaces) προς τους servers του rack και τις διεπαφές βόρειας κατεύθυνσης (northbound interfaces) προς τα υψηλότερα επίπεδα του δέντρου. Αν υποθέσουμε ότι σε ένα datacenter υπάρχουν  $N$  κεντρικοί υπολογιστές και  $k$  μεταγωγείς πακέτων, τότε το βάθος της τοπολογίας fat-tree υπολογίζεται από τον μαθηματικό τύπο:

$$D = \lceil (\log N - 1) / (\log k - 1) \rceil$$

Ως εκ τούτου, η πιθανή αύξηση του αριθμού των endhost (servers) σε μια τέτοια αρχιτεκτονική οδηγεί και στην απαίτηση εγκατάστασης ενός επιπλέον επιπέδου στο fat-tree μόλις καλυφθεί ένας αριθμός από servers. Σήμερα, τα κέντρα δεδομένων διαθέτουν δεκάδες έως εκατοντάδες χιλιάδες διακομιστές και απαιτούνται τρία ή τέσσερα επίπεδα fat-tree για

να επιτευχθεί πλήρης αξιοποίηση στο εύρος ζώνης σε όλες τις κατευθύνσεις (full bisection bandwidth). Ως εκ τούτου, παρόλο που το κόστος της Clos προσέγγισης είναι χαμηλότερο από την παραδοσιακή προσέγγιση των fat-tree, εξακολουθεί να κλιμακώνεται υπερ-γραμμικά ανάλογα με τον αριθμό των επιπρόσθετων endhosts. Επιπλέον, η καλωδίωση αυτού του τεράστιου αριθμού διακοπών γίνεται αρκετά δύσκολη και επιρρεπής σε σφάλματα κατά την εγκατάσταση και τη συντήρηση.

Επίσης, η χρήση μεγάλου αριθμού ηλεκτρικών διακοπών/μεταγωγέων πακέτου αυξάνει σημαντικά την κατανάλωση ενέργειας ολόκληρου του συστήματος. Για παράδειγμα οι μεταγωγείς 10 GbE 64 ports καταναλώνουν 150 έως 350 Watt. Εδώ πρέπει να σημειώσουμε ότι περίπου το 90% αυτής της κατανάλωσης ενέργειας είναι ανεξάρτητο από το φορτίο και επομένως η εξοικονόμηση είναι αδύνατο να προέλθει από οποιαδήποτε μέθοδο εξισορρόπησης ή προγραμματισμού φορτίου.

Τέλος, η δυνατότητα αναβάθμισης (upgradability) είναι ένα μεγάλο πρόβλημα που υπάρχει στις fat-tree αρχιτεκτονικές. Κι αυτό οφείλεται σε δύο εγγενείς λόγους:

- (i) Η προσθήκη περισσότερων rack-servers απαιτεί τη σύνδεση πολλών ελεύθερων θυρών και διάσπαρτων μεταγωγέων, με την προϋπόθεση ότι υπάρχουν διαθέσιμες ελεύθερες θύρες, δηλαδή δεν έχουμε φτάσει το όριο διακομιστή για το συγκεκριμένο βάθος του δέντρου, αλλά μόλις επιτευχθεί το όριο, απαιτείται ένας τεράστιος αριθμός νέων μεταγωγέων για να συμπληρωθεί το νέο επίπεδο,
- (ii) Η αναβάθμιση του ρυθμού διασύνδεσης των διακομιστών απαιτεί ένα πλήρες νέο δίκτυο fat-tree και στις περισσότερες περιπτώσεις οι προηγούμενοι διακόπτες δεν μπορούν να επαναχρησιμοποιηθούν.

Τα δίκτυα τύπου fat-tree τις περισσότερες φορές υπό-χρησιμοποιούνται [13]. Ως αποτέλεσμα, δεν απαιτείται πλήρης διαμοιρασμός του εύρους ζώνης, αφού σπάνια συμβαίνει όλοι οι διακομιστές να μιλούν ταυτόχρονα σε πλήρη ταχύτητα. Αυτό επιλύεται εν μέρει με τη δημιουργία ενός over-subscribed δικτύου fat-tree, δηλαδή ενός δέντρου με μικρότερο εύρος ζώνης στα υψηλότερα επίπεδα, το οποίο δεν παρέχει πλήρη διαμοιρασμό του εύρους ζώνης αλλά είναι φθηνότερο. Ακόμη και οι τοπολογίες over-subscribed fat-tree χρησιμοποιούνται λιγότερο κατά μέσο όρο [13]. Πιο συγκεκριμένα, η μέση χρήση για ένα over-subscribed fat-tree δίκτυο είναι λιγότερη από 20%, ενώ υπάρχουν ορισμένες στιγμές όπου δημιουργείται έντονη συμφόρηση με τη μορφή hotspot με αποτέλεσμα να χάνονται πακέτα. Το πρόβλημα είναι, τελικά, η μη-δυναμική κατανομή του διαθέσιμου (μειωμένου) εύρους ζώνης στους διακομιστές που δεν διευκολύνει την ορθολογική κατανομή των διαθέσιμων πόρων μεταξύ των endhost ανάλογα με τις ανάγκες του δικτύου.

Από την άλλη, η καθυστέρηση (latency) εξαρτάται από το φορτίο του συστήματος, καθώς η συμφόρηση προσθέτει αλληπάλληλες καθυστερήσεις και μπορεί να οδηγήσει σε απώλειες πακέτων που αντιμετωπίζονται από τα υψηλότερα επίπεδα (TCP) ή από την εκάστοτε

εφαρμογή. Ακόμη και όταν είναι εγγυημένη η λειτουργία χωρίς απώλειες (π.χ. με χρήση της Infiniband ή του Ethernet Datacentre Bridging), η συμφόρηση εξακολουθεί να αυξάνει την καθυστέρηση. Μπορούμε ωστόσο να μετρήσουμε τη μεγαλύτερη καθυστέρηση διαδρομής ενός κενού δικτύου (σε κατάσταση μηδενικού φορτίου), όπου η καθυστέρηση στην ουρά είναι μηδέν, ενώ η καθυστέρηση διάδοσης αγνοείται. Τα τυπικά μεγέθη των καθυστερήσεων του Ethernet switch είναι μεταξύ 500 nsec και 1,5 μsec. Σε ένα δίκτυο με ελαφρύ φορτίο η τάξη της καθυστέρησης είναι κάποια εκατοντάδες μsec ή ελάχιστα msec.

Εν ολίγοις, οι κύριοι περιορισμοί των σημερινών αρχιτεκτονικών κέντρων δεδομένων είναι:

- Υπερ-γραμμική κλιμάκωση των διαστάσεων
- Υψηλή κατανάλωση ενέργειας ανεξάρτητα από το φορτίο
- Τεράστιος αριθμός καλωδίων (cable spaghetti)
- Σπατάλη capex και opex αφού κατά μέσο όρο το δίκτυο υποχρησιμοποιείται
  - Η πλήρης fat-tree τοπολογία είναι πολύ ακριβή
  - Η τοπολογία over-subscribed fat-tree (και εδώ το δίκτυο υποχρησιμοποιείται) μπορεί να παρουσιάσει έντονα προβλήματα συμφόρησης (hotspot)
- Μη-δυναμική κατανομή εύρους ζώνης
- Μεγάλοι χρόνοι καθυστέρησης της μετάδοσης της κυκλοφορίας στον οριζόντιο άξονα east-west (αφού τα πακέτα πρέπει να «ταξιδέψουν» μέσω του κατακόρυφου άξονα south-north)

### 1.1.2 Η πρόκληση του διαχωρισμού των πόρων (disaggregation)

Οι παραπάνω περιορισμοί επιδεινώνονται περαιτέρω από την εφαρμογή της τεχνικής Διαχωρισμού των Πόρων (Resources disaggregation), η οποία εμφανίζει πλέον ραγδαία δυναμική. Μέχρι στιγμής, στο συμβατικό μοντέλο κέντρων δεδομένων, οι υπολογιστικοί πόροι, οι μνήμες, τα μέσα αποθήκευσης και επικοινωνίας είναι σταθερά για τους servers που απαρτίζουν τα κέντρα δεδομένων. Πράγματι, κάθε server αποτελείται από έναν σταθερό συνδυασμό υπολογιστικών πόρων, πόρων μνήμης, αποθήκευσης και διεπαφών επικοινωνίας που είναι όλοι ενσωματωμένοι ή «συγκεντρωμένοι» στο ίδιο μηχάνημα. Η βασική ιδέα του διαχωρισμού των πόρων είναι η κοινή χρήση αυτών μεταξύ των rack του datacenter και η χρήση τους κατ' απαίτηση (on-demand), δηλαδή δυναμικά ανάλογα με τις ανάγκες.

Υπάρχουν πολλαπλά οφέλη από τη μετάβαση σε κέντρα δεδομένων με καταναμημένους τους διαθέσιμους πόρους. Αυτές οι υποδομές που αποτελούνται στην πραγματικότητα από επιμέρους «κομμάτια» (modularity) οδηγούν σε πιο αποτελεσματική λειτουργία και

βελτιωμένη απόδοση συνολικά. Αυτή η φυσική αποσύνδεση των πόρων επιτρέπει πιο λεπτομερή και ελεγχόμενη διάθεση και αξιοποίησή τους, ενώ ταυτόχρονα επιτυγχάνεται μεγαλύτερο ποσοστό χρήσης τους με πολυπλεξία των διαθέσιμων πόρων [14]. Επιπλέον, αυτή η «κομματιαστή» δομή επιτρέπει την επεκτασιμότητα του δικτύου ανεξάρτητα για κάθε «κομμάτι» πόρων, καθώς κάθε διαφορετικός τύπος πόρων ακολουθεί διαφορετικές τάσεις και διέπεται από διαφορετικούς περιορισμούς. Από αυτή την άποψη, οι ιδιοκτήτες/διαχειριστές αυτών των σύγχρονων δικτύων Κέντρων Δεδομένων είναι σε θέση να προσαρμόσουν εύκολα κάθε ξεχωριστό «κομμάτι» σε οποιαδήποτε νέα τεχνολογία ή νέα εκδοχή ανεξάρτητα από τα υπόλοιπα «κομμάτια» των υπόλοιπων πόρων. Για παράδειγμα η αναβάθμιση σε μια νέα γενιά συστήματος αποθήκευσης αφήνει ανεπηρέαστα όλα τα στοιχεία υπολογιστικής ισχύος, μνήμης και δικτύου και έτσι εξοικονομείται περιττό κόστος. Έτσι δίνεται και η δυνατότητα για επιπλέον ευελιξία στην ανάπτυξη καινοτόμων εξαρτημάτων.

Η έννοια του διαχωρισμού των διαθέσιμων πόρων υποστηρίχθηκε κυρίως από τις μεγάλες εταιρείες/παρόχους, και οδήγησε στην ίδρυση του Open Compute Project (OCP) που στοχεύει στην οικοδόμηση μιας ευρείας βιομηχανικής συναίνεσης. Η προτεινόμενη αρχιτεκτονική, που ονομάζεται Disaggregated Rack-Scale Server (DRS), είναι βελτιστοποιημένη για να εξυπηρετεί μεγάλο χώρο διευθύνσεων, βάσεις δεδομένων με ενσωματωμένη μνήμη και ανάλυση δεδομένων (data analytics). Ο στόχος του OCP/DRS είναι ο διαχωρισμός των πόρων υπολογιστικής ισχύος, αποθήκευσης και επικοινωνίας μέσα στο rack και η διασύνδεση μεταξύ τους με καταναμημένες λειτουργίες μεταγωγής [11]. Το OCP αναμένει μείωση του κόστους κατά 24% και αύξηση της απόδοσης κατά περίπου 38% με αυτό το νέο παράδειγμα διαχωρισμού μέσα στο rack [15]. Παρόλο που η ιδέα του διαχωρισμού των πόρων έχει αρχίσει να υλοποιείται εδώ και κάποια χρόνια [16] είναι βέβαιο ότι γρήγορα θα κυριαρχεί παντού.

Παρά τα μοναδικά πλεονεκτήματα που προσφέρει η υιοθέτηση της ιδέας του διαχωρισμού των πόρων όσον αφορά την εξομοίωση (virtualization), τη βέλτιστη κατανομή των πόρων και τη δυνατότητα αναβάθμισης των υποδομών, συνοδεύεται από ένα σημαντικό εμπόδιο: ο διαχωρισμός των πόρων του συστήματος προκαλεί υπερβολικές απαιτήσεις στο δίκτυο που τους διασυνδέει. Αυτό έχει σαν αποτέλεσμα, οι διασυνδέσεις του δικτύου να έρχονται αντιμέτωπες με την πρόκληση να ανταπεξέλθουν επιτυχώς στις αυξανόμενες απαιτήσεις για υψηλό εύρος ζώνης και ελάχιστη καθυστέρηση κατά μήκος όλης της απόστασης των καταναμημένων πόρων στο κέντρο δεδομένων.

Οι τρέχουσες υλοποιήσεις που βασίζονται στις συμβατικές fat-tree αρχιτεκτονικές και στους σημερινούς διακόπτες/μεταγωγείς, απλά δεν μπορούν να επεκταθούν εύκολα για να υποστηρίξουν αυτές τις απαιτήσεις αυξημένων δεδομένων. Σε αυτό το πλαίσιο, η φωτονική τεχνολογία μπορεί να αποτελέσει μια πιθανή λύση καθώς προσφέρεται για μεταφορά δεδομένων υψηλού εύρους ζώνης [17]. Όσον αφορά στην αποτελεσματική μεταγωγή και

δρομολόγηση αυτού του τεράστιου όγκου πληροφοριών, η φωτονική εμφανίζεται και πάλι ως η τεχνολογία που μπορεί να δώσει λύσεις. Η χρήση του οπτικού μέσου όχι μόνο ως φορέα μετάδοσης αλλά και ως τρόπου μεταγωγής μπορεί να αποσυμφορήσει το φορτίο δεδομένων από τους ηλεκτρονικούς μεταγωγείς που απαιτούν αρκετή υπολογιστική ισχύ και επίσης να επιτρέψει την ανάπτυξη πιο επίπεδων αρχιτεκτονικών δικτύου.

Όπως συμβαίνει και στα οπτικά τηλεπικοινωνιακά δίκτυα, η ιδέα δεν είναι να αντικαταστήσουμε τα ηλεκτρονικά μέρη με οπτικά/φωτονικά, αλλά να τα χρησιμοποιήσουμε «συνεργατικά», ώστε να εκμεταλλευτούμε το καλύτερο και των δύο αυτών κόσμων. Αυτή η ιδέα έχει τροφοδοτήσει μια σειρά από ερευνητικές δράσεις σε υβριδικά ηλεκτρο-οπτικά δίκτυα που θεωρούνται ένα βιώσιμο μονοπάτι μετάβασης προκειμένου τα δίκτυα κέντρων δεδομένων να ανταποκριθούν στις αυξανόμενες απαιτήσεις.

### **1.1.3 Αρχιτεκτονικές οπτικής μεταγωγής για Δίκτυα Κέντρων Δεδομένων**

Η οπτική μεταγωγή μπορεί να αξιοποιηθεί σε περιπτώσεις συγκεντρωμένης κίνησης μεταξύ των racks ή ομάδων από rack, αντικαθιστώντας εν μέρει ή εξ ολοκλήρου τα υψηλότερα επίπεδα των fat-tree των ηλεκτρονικών δικτύων [18], [19], [20]. Έχουν προταθεί αρχιτεκτονικές που αξιοποιούν κάποια σημαντικά χαρακτηριστικά που έχουν οι οπτικοί μεταγωγείς, όπως είναι η διαφάνεια ως προς το πρωτόκολλο και τον ρυθμό μετάδοσης (protocol & datarate transparency), το σχετικά χαμηλό κόστος ανά θύρα και η δυνατότητα επαναδιαμόρφωσης (reconfigurability). Η επαναδιαμόρφωση του οπτικού μεταγωγέα είναι ζωτικής σημασίας για τη λειτουργία του δικτύου, καθώς μπορεί να βελτιώσει την αποτελεσματικότητα και να μειώσει το κόστος παρέχοντας on-demand διασυνδέσεις δυναμικού εύρους ζώνης, αποφεύγοντας την υπερβολική ή αχρεία στατική κατανομή εύρους ζώνης (υπερπληθών ή μη) fat-tree δικτύων. Εμπορικά διαθέσιμες συσκευές (Commercial Off-The-Shelf - COTS) όπως MEMS, προσαρμοζόμενα λείζερ και οπτικά φίλτρα AWG έχουν διερευνηθεί ως λύσεις ενός τέτοιου ρεαλιστικού σεναρίου.

Μια κατηγορία δικτύων κέντρων δεδομένων με οπτική μεταγωγή βασίζεται σε μεταγωγείς Micro Electro-Mechanical Systems (MEMS), λόγω της ωριμότητας της τεχνολογίας τους. Οι μεταγωγείς MEMS έχουν αρκετά μεγάλους χρόνους επαναδιαμόρφωσης, που συνήθως κυμαίνονται από δεκάδες έως εκατοντάδες χιλιοστά του δευτερολέπτου, καθιστώντας τους κατάλληλους για μεταγωγές κυκλώματος. Για να εξυπηρετηθεί επίσης η κίνηση με μεταγωγή πακέτων, το οπτικό δίκτυο σε αυτές τις περιπτώσεις χρησιμοποιείται παράλληλα με ένα ηλεκτρικό δίκτυο μεταγωγής πακέτων. Έτσι, οι ροές μεγάλης διάρκειας διαχειρίζονται στο δίκτυο οπτικής μεταγωγής, ενώ μικρότερες ροές εξυπηρετούνται από το ηλεκτρικό δίκτυο μεταγωγής πακέτων. Αυτή η προσέγγιση υβριδικού ηλεκτρονικού-οπτικού δικτύου που βασίζεται σε MEMS ακολουθείται για παράδειγμα στα δίκτυα Helios [18], Calient [21] και REAToR [22].

Η τεχνική μεταγωγής οπτικού κυκλώματος, ως τεχνική δρομολόγησης, παρέχει ουσιαστικά απεριόριστο εύρος ζώνης που μεταβάλλεται χωρίς την ανάγκη αναβάθμισης του εξοπλισμού ενώ παράλληλα οι ταχύτητες του δικτύου αυξάνονται. Από την άλλη πλευρά, ένα σημαντικό μέρος της κίνησης στο κέντρο δεδομένων είναι βραχύβια και δεν μπορεί να επωφεληθεί από τα πλεονεκτήματα ενός δικτύου οπτικής μεταγωγής. Επιπλέον, σε ένα τέτοιο πλαίσιο απαιτείται ταξινόμηση της κίνησης που όμως είναι εξαιρετικά απαιτητική με τη σειρά της αφού θα πρέπει να συνοδεύεται από ταυτόχρονη παρακολούθηση και πρόβλεψη της κίνησης μέσα σε ένα εξαιρετικά μεγάλο δίκτυο δεδομένων. Σε αυτό το σκέλος συμβάλει το επίπεδο ελέγχου και «ενορχήστρωσης» του δικτύου που αναλαμβάνει τον χειρισμό του δικτύου, μια διαδικασία που προσθέτει σημαντικές καθυστερήσεις [23] σε κλίμακα των δευτερολέπτων. Ένα από τα σημεία συμφόρησης στο πεδίο του ελέγχου είναι ότι η επαναδιαμόρφωση του δικτύου, περιλαμβάνει εκτός από την ρύθμιση των MEMS μεταγωγέων, την ενημέρωση του κεντρικού υπολογιστή ή των διακοπών σχετικά με τον τρόπο διαχωρισμού της κυκλοφορίας σε επιμέρους κυκλώματα. Τέλος, δεδομένου ότι οι διαστάσεις των MEMS είναι αρκετά περιορισμένη (διατίθεται μεταγωγείς έως και 320 θυρών στο εμπόριο) και η κατασκευή μεταγωγέων με περισσότερες θύρες με συνδυασμών μικρότερων μεταγωγέων είναι πολύ περίπλοκη (λόγω απωλειών και προβλημάτων συγχρονισμού), η υβριδική λύση που βασίζεται σε MEMS παρουσιάζει σημαντικά προβλήματα όσον αφορά την επεκτασιμότητα της.

Στην προσπάθεια μείωσης των χρόνων επαναρρύθμισης του οπτικού δικτύου, έχουν προταθεί οι τεχνικές μεταγωγής μήκους κύματος που μπορούν να αποτελέσουν μια βιώσιμη επιλογή στην διαδικασία οπτικής δρομολόγησης πακέτων. Τέτοιες τεχνικές βασίζονται στη δυνατότητα των προσαρμοζόμενων λείζερ να συντονίζονται πολύ γρήγορα καθώς και στον συνδυασμό τους με οπτικά στοιχεία επιλογής μήκους κύματος, όπως είναι οι οπτικοί δρομολογητές πλέγματος κυματοδηγών σε διάταξη (Arrayed Waveguide Grating Routers - AWGR), οι οποίοι μπορούν να δρομολογήσουν τα εισερχόμενα οπτικά δεδομένα στην αντίστοιχη έξοδο σύμφωνα με το μήκος κύματος τους.

Επιπλέον, οι υλοποιήσεις που χρησιμοποιούν μεταγωγή μήκους κύματος με AWGR και άλλα παθητικά οπτικά στοιχεία είναι ανεκτικές σε παρεμβολές (crosstalk) ενώ δεν χειροτερεύουν περεταίρω τα σήματα κατά τη διαδικασία της μεταγωγής [24]. Αρκετές ερευνητικές ομάδες έχουν διερευνήσει αυτήν την τεχνική μεταγωγής μήκους κύματος σε διαφορετικές υλοποιήσεις, όπως είναι τα DOS [25], LIONS [26], το Petabit [27] και το IRIS [28], [29], [30]. Παρά τα πλεονεκτήματα που παρουσιάζουν αυτές οι αρχιτεκτονικές και αφορούν στην ταχύτητα, την «κομματιαστή» τους δομή και τη χαμηλή πιθανότητα υπερσυγκέντρωσης κίνησης σε ένα δίκτυο κέντρων δεδομένων, αντιμετωπίζουν σημαντικά προβλήματα επεκτασιμότητας. Ο συνολικός αριθμός μηκών κύματος για τις οπτικές ζεύξεις είναι περιορισμένος (περίπου 80 στη ζώνη C), γεγονός που εμποδίζει την πιθανή επέκταση του δικτύου σε δίκτυα κέντρων δεδομένων μεγάλης κλίμακας.

Μια άλλη υβριδική προσέγγιση αρχιτεκτονικής κέντρων δεδομένων που στοχεύει να μειώσει τους μεγάλους χρόνους επαναδιαμόρφωσης των οπτικών μεταγωγέων MEMS παρουσιάζεται στο έργο Mordia [31]. Πιο συγκεκριμένα, προτείνεται η τεχνική Πολυπλεξίας διαίρεσης μήκους κύματος (Wavelength Division Multiplexing – WDM) με γρήγορους διακόπτες/μεταγωγείς επιλογής μήκους κύματος (Wavelength Selective Switch - WSS) για την δρομολόγηση εντός του δικτύου. Τα WSS που χρησιμοποιούνται παρουσιάζουν χρόνους μεταγωγής της τάξης των 10 μs, ο οποίος είναι σημαντικά χαμηλότερος από την ταχύτητα μεταγωγής των MEMS, αλλά εξακολουθεί να μην είναι απολύτως συμβατός με τον τρόπο που πρέπει να διαχειριστεί η κίνηση πακέτων τύπου Ethernet. Εκτός από την αρχιτεκτονική και την υλοποίηση του φυσικού επιπέδου, στο έργο αυτό διερευνώνται αλγόριθμοι που εξασφαλίζουν γρήγορη επαναδιαμόρφωση της υποδομής του δικτύου. Αν και οι προτεινόμενοι αλγόριθμοι είναι 2-3 τάξεις μεγέθους ταχύτεροι από τις παραδοσιακές προσεγγίσεις, εξακολουθούν να μην μπορούν να προσαρμοστούν και να εφαρμοστούν σε τεράστια κέντρα δεδομένων που μπορεί να αποτελούνται από χιλιάδες rack. Οι κύριοι περιορισμοί και εδώ, παραμένουν η επεκτασιμότητα και το κόστος. Επίσης, τόσο η αρχιτεκτονική όσο και οι λειτουργίες του επιπέδου ελέγχου δεν έχουν δοκιμαστεί πιλοτικά σε κάποιο κέντρο δεδομένων ρεαλιστικού μεγέθους [32].

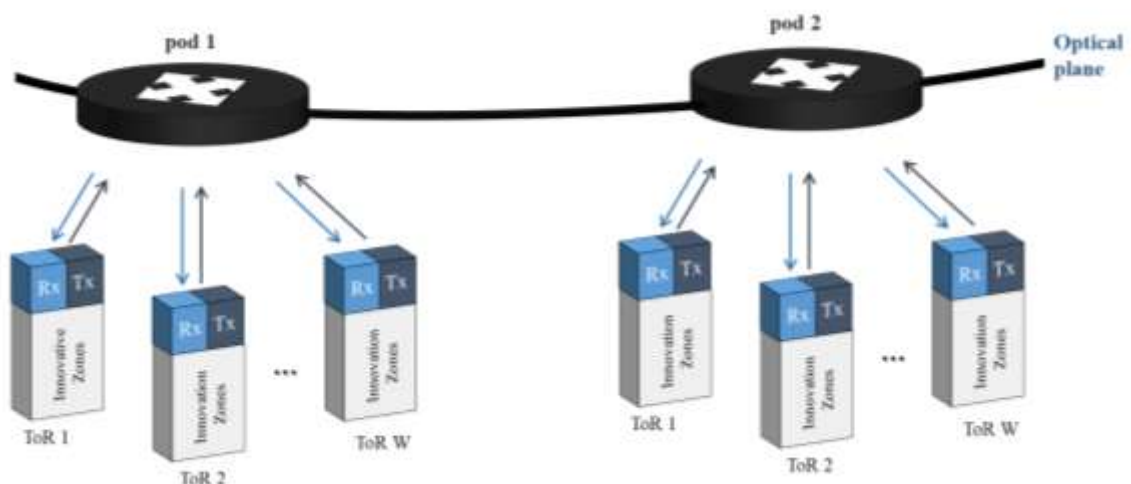
Τέλος, μια επίσης εναλλακτική υβριδική ηλεκτρο-οπτική προσέγγιση έχει παρουσιαστεί και από το έργο Lightness [33] σε μια προσπάθεια να εξαλειφθούν εντελώς οι ηλεκτρικοί μεταγωγείς από το κέντρο δεδομένων. Προτείνεται ένα δίκτυο κέντρων δεδομένων που μπορεί να διαχειριστεί με οπτικό τρόπο τόσο ροές πακέτων όσο και κυκλωμάτων. Επιπλέον, υιοθετείται ένα επίπεδο δεδομένων (data-plane) που βασίζεται στην οπτική τεχνολογία, το οποίο υλοποιείται σε μια επίπεδη αρχιτεκτονική που ενσωματώνει τόσο τις τεχνικές μεταγωγής οπτικών πακέτων (Optical Packet Switching - OPS) όσο και τις τεχνικές μεταγωγής οπτικών κυκλωμάτων (Optical Circuit Switching - OCS). Οι μεταγωγείς πακέτων OPS χρησιμοποιούνται για τις ροές μικρής διάρκειας με απαιτήσεις χαμηλού χρόνου καθυστέρησης, ενώ οι μεταγωγείς κυκλωμάτων OCS για ροές μεγάλης διάρκειας. Οι διακομιστές (servers) διασυνδέονται με τα υβριδικά δίκτυα κέντρων δεδομένων OPS/OCS μέσω ηλεκτρικών ToR μεταγωγέων, οι οποίοι συγκεντρώνουν την κίνηση και την ταξινομούν ανάλογα με τη διάρκεια που έχει. Ο ToR μεταγωγέας είναι ηλεκτρικός (με χρήση FPGA) και έχει σχεδιαστεί έτσι ώστε να παρέχει χωρητικότητα 100 Gb/s και χαμηλή καθυστέρηση στο υβριδικό OCS/OPS. Ο OPS μεταγωγέας υποστηρίζει WDM λειτουργία και βασίζεται σε διακόπτες επιλογής μήκους κύματος που υλοποιούνται με AWG ακολουθούμενα από μεγάλες συστοιχίες Οπτικών ενισχυτών ημιαγωγών (Semiconductor Optical Amplifier - SOA) για την δρομολόγηση πακέτων μεταξύ διαφορετικών ToR.

Αυτή η ιδέα μεταγωγής που βασίζεται σε SOAs έχει διερευνηθεί και στα πλαίσια των τηλεπικοινωνιακών δικτύων OPS μεταγωγής [34], ωστόσο εξακολουθεί να παρουσιάζει

σημαντικά μειονεκτήματα, αφού αφενός παρουσιάζει σχετικά υψηλή κατανάλωση ενέργειας και αρκετά μεγάλη απαγωγή της ισχύος, και αφετέρου υπάρχει σημαντική έλλειψη στην αλυσίδα ανεφοδιασμού με τα απαραίτητα εξαρτήματα. Η μεταγωγή οπτικών κυκλωμάτων (OCS) πραγματοποιείται μέσω κόμβων Architecture on Demand (AoD) όπου ένα οπτικό backplane με μεγάλο αριθμό από MEMS μεταγωγείς συνδέεται με πολλές μονάδες επεξεργασίας σήματος καθώς και με τις εισόδους/εξόδους του κόμβου προσφέροντας ευελιξία, αφού τα συστήματα δεν είναι συνδεδεμένα στατικά με ενσύρματο τρόπο αλλά μπορούν να διασυνδεθούν μεταξύ τους με τυχαίο τρόπο. Αυτό όμως μπορεί να οδηγήσει στην δέσμευση περιττού οπτικού εξοπλισμού που αυξάνει το συνολικό κόστος κατασκευής και τα έξοδα διαχείρισης του δικτύου. Τέλος, δεδομένου ότι οι διαστάσεις των MEMS δεν μπορούν να επεκταθούν εύκολα, ένα τεράστιο πρόβλημα στην AoD-OCS προσέγγιση είναι η περιορισμένη επεκτασιμότητα του δικτύου.

## 1.2 Εισαγωγή στην υβριδική αρχιτεκτονική

Η προτεινόμενη αρχιτεκτονική δικτύου για Κέντρα Δεδομένων απεικονίζεται στην Εικόνα 3 και υλοποιήθηκε στα πλαίσια του ευρωπαϊκού έργου NEPHELE. Στο έργο εκτός από το Εθνικό Μετσόβιο Πολυτεχνείο μέσω του ΕΦΕ, συμμετείχαν σημαντικές εταιρείες από το εξωτερικό αλλά και Πανεπιστήμια και Ερευνητικά Ινστιτούτα.



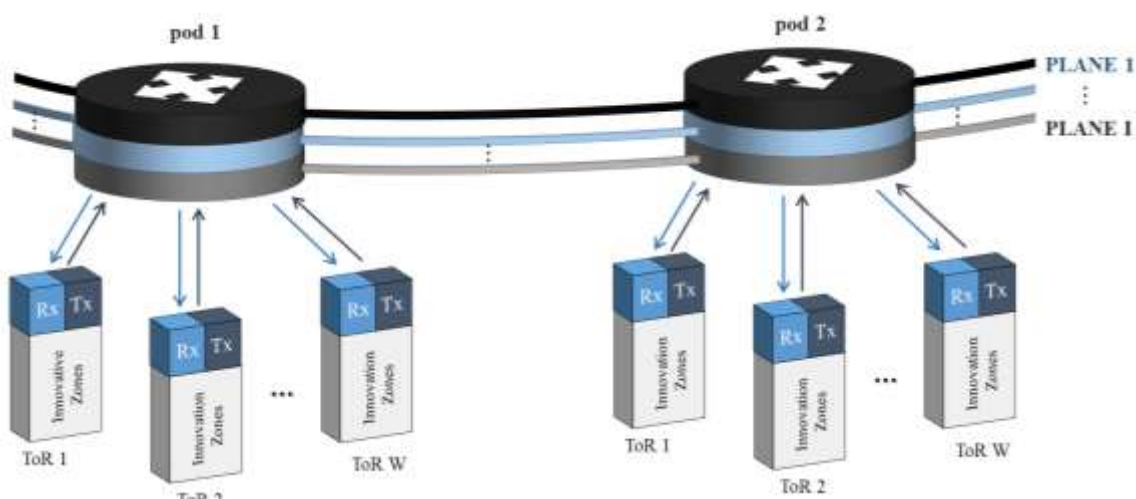
*Εικόνα 3: Σχηματικό της προτεινόμενης υβριδικής αρχιτεκτονικής*

Το κύριο δομικό στοιχείο της αρχιτεκτονικής είναι το pod, που φιλοξενεί έναν αριθμό από racks, που με τη σειρά τους εξυπηρετούν μερικές χιλιάδες από καταναμημένους πόρους (π.χ. χώρους αποθήκευσης, υπολογιστική ισχύ κτλ) που ονομάζονται «ζώνες καινοτομίας». Ως εκ τούτου, το pod είναι ουσιαστικά ένα κέντρο δεδομένων μικρής κλίμακας. Κάθε rack είναι υπό τη διαχείριση ενός top-of-rack μεταγωγέα (ToR switch), ενώ όλοι οι ToR switch



διασυνδέονται με το pod switch με οπτικό τρόπο σε μια τοπολογία αστέρα, χρησιμοποιώντας μία θύρα ανά ToR.

Κάθε ToR switch είναι εξοπλισμένο με ένα προσαρμοζόμενο οπτικό πομπό και έναν οπτικό δέκτη που υποστηρίζει την λειτουργία πακέτων. Για την κίνηση που προορίζεται εντός του ίδιου pod (intra-pod), η οπτική δρομολόγηση πραγματοποιείται κυρίως μέσω παθητικών στοιχείων οπτικού φιλτραρίσματος. Παράλληλα, για την επεκτασιμότητα του δικτύου, πολλαπλά pod μπορούν να διασυνδεθούν με οπτικές ίνες σε μια τοπολογία δακτυλίου. Κάθε οπτικός δακτύλιος μεταφέρει πολυπλεγμένη κίνηση με διαίρεση μήκους κύματος (WDM) και αποτελείται από πολλαπλές οπτικές ίνες για την παροχή της απαραίτητης χωρητικότητας μεταξύ των pods. Η επικοινωνία μεταξύ διακομιστών που βρίσκονται μεταξύ διαφορετικών pods (inter-pod) είναι ένας συνδυασμός μεταγωγής μήκους κύματος (wavelength switching) αλλά και χώρου (space switching), επιτρέποντας την επαναχρησιμοποίηση των μηκών κύματος μεταξύ των pods, και έτσι επιτυγχάνεται επεκτασιμότητα του δικτύου πέρα από την τυπική, συγκριτικά με τον αριθμό των μηκών κύματος σε πυκνά συστήματα WDM (Dense-WDM).



**Εικόνα 4:** *Επέκταση του δικτύου με την προσθήκη επιπλέον οπτικών επιπέδων*

Η είσοδος ή έξοδος της κίνησης προς και από τον οπτικό δακτύλιο βασίζεται στο μήκος κύματος και η τεχνική δρομολόγησης ονομάζεται add/drop multiplexing. Ως εκ τούτου, παρά την τοπολογία δακτυλίου σε φυσικό επίπεδο, η λογική της τοπολογίας του δικτύου είναι επί της ουσίας τοπολογία πλέγματος (mesh topology). Στο πλαίσιο αυτό, το σύνολο ενός οπτικού δακτυλίου μαζί με τους αντίστοιχους pod και ToR switches ονομάζεται οπτικό επίπεδο (optical plane). Η αύξηση της χωρητικότητας του δικτύου επιτυγχάνεται με την προσθήκη επιπλέον ανεξάρτητων οπτικών επιπέδων (optical planes). Για να γίνει αυτό, απαιτείται η εγκατάσταση πρόσθετων pod switches και η διασύνδεσή τους μέσω νέων

οπτικών δακτυλίων, καθώς και συμπλήρωση πρόσθετων θυρών στους ToR switches για τη διασύνδεση με τους νέους pod switches, όπως φαίνεται στην Εικόνα 4.

Η προσθήκη νέων οπτικών επιπέδων δεν επηρεάζει τα υπάρχοντα, διασφαλίζοντας έτσι την επεκτασιμότητα και επιτρέποντας την ανάπτυξη και επέκταση ανάλογα με τις τρέχουσες ή τις μελλοντικές απαιτήσεις. Οι τιμές αναφοράς των παραμέτρων της προτεινόμενης υβριδικής αρχιτεκτονικής συνοψίζονται στον Πίνακα 1.

**Πίνακας 1:** Διαστάσεις του επιπέδου δεδομένων της προτεινόμενης αρχιτεκτονικής δικτύου

Παράμετρος	Σημασία	Τυπική τιμή
<b>W</b>	Ο αριθμός των racks και των ToR switches σε κάθε pod. Επίσης, ο αριθμός των μηκών κύματος στο σύστημα	80
<b>R</b>	Ο αριθμός των οπτικών δακτυλίων για κάθε οπτικό επίπεδο (fiber rings per optical plane)	20
<b>P</b>	Ο αριθμός των pods	20
<b>I</b>	Ο αριθμός των οπτικών επιπέδων (optical planes)	20

Το επίπεδο δεδομένων (dataplane) της αρχιτεκτονικής λειτουργεί με βάση την τεχνική TDMA με χρονοσχισμές ή αλλιώς timeslots, όπου τα «slots» είναι τα χρονικά τμήματα στα οποία είναι δυνατή η επικοινωνία μεταξύ των racks (rack-to-rack). Τα slots (και επομένως οι πόροι δικτύου) μπορούν να διατεθούν δυναμικά στην επικοινωνία μεταξύ των rack και έτσι το δίκτυο μπορεί να πετύχει σχεδόν πλήρη χρήση της χωρητικότητάς του, οδηγώντας σε μείωση της κατανάλωσης της ενέργειας αλλά και κόστους.

Η λειτουργία σε slots και η επεκτασιμότητα του προτεινόμενου δικτύου με χρήση επιπλέον οπτικών επιπέδων μπορεί να αναβαθμίσει σημαντικά τις επιδόσεις των οπτικών κέντρων δεδομένων, ενώ παράλληλα βασίζεται σε φωτονικά στοιχεία που ήδη είναι εμπορικά διαθέσιμα [22] [35]. Σε αντίθεση με τις προσεγγίσεις που βασίζονται στην κατανομή ελαστικού φάσματος [8], η προσέγγιση TDMA παρέχει δυναμική ανάθεση της χωρητικότητας δικτύου χωρίς την ανάγκη περίπλοκου flex-grid υλικού, που θα αύξανε δραματικά το κόστος κατασκευής. Για πολύ δυναμικά σενάρια κυκλοφορίας που κυριαρχούνται από μικρές ροές της λεγόμενες «mice flows», εξετάζεται μια υβριδική ηλεκτρονική-οπτική υλοποίηση στην προτεινόμενη αρχιτεκτονική, με δύο δίκτυα να υλοποιούνται στο επίπεδο των ToR switches.

Η τοπολογία είναι ένα δίκτυο δύο επιπέδων: το πρώτο επίπεδο περιλαμβάνει τους ToR switches και πάνω από αυτούς υπάρχει ένα ενιαίο επίπεδο pod switches. Για την υποστήριξη περισσότερων servers, το δίκτυο επεκτείνεται στην οριζόντια κατεύθυνση («east-west»), με

την προσθήκη επιπλέον pods, εξυπηρετώντας πολύ καλύτερα την «east-west» κίνηση που κυριαρχεί στα κέντρα δεδομένων. Έτσι, κατά μία έννοια, το δίκτυο είναι επίπεδο, σε σύγκριση με τα κλασικού τύπου fat-tree δίκτυα που δρομολογούν την κυκλοφορία σε διάφορα επίπεδα, ο αριθμός των οποίων εξαρτάται από τον αριθμό των servers. Αξίζει να σημειωθεί ότι ο απαιτούμενος εξοπλισμός του δικτύου αυξάνεται γραμμικά στην προτεινόμενη αρχιτεκτονική, ενώ σε fat-tree αρχιτεκτονικές απαιτείται σε όλα τα επίπεδα η προσθήκη επιπλέον switches και από ένα σημείο και μετά την προσθήκη νέου επιπέδου, οδηγώντας έτσι σε υπεργραμμική αύξηση του αριθμού των servers.

### 1.3 Η δομή και τα στοιχεία της προτεινόμενης αρχιτεκτονικής

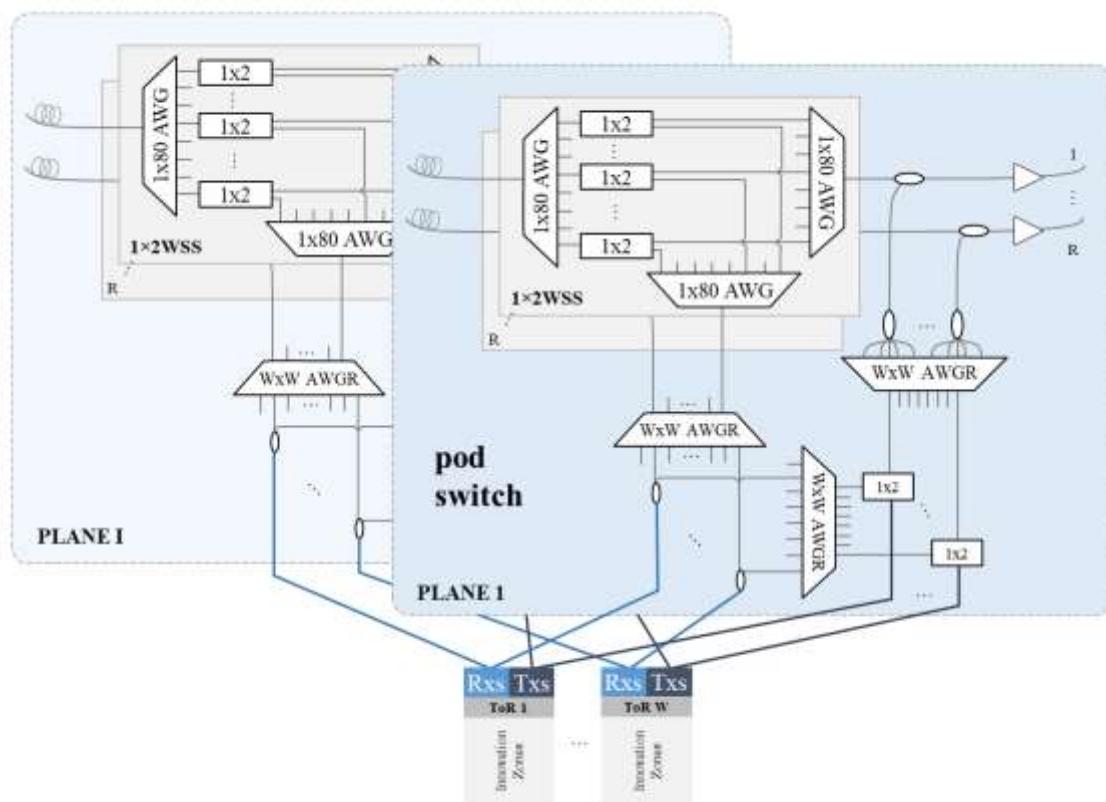
Η προτεινόμενη υβριδική αρχιτεκτονική που παρουσιάζεται στην παρούσα διατριβή συνδυάζει, όπως αναφέρθηκε και προηγουμένως, τα πλεονεκτήματα των τεχνικών της οπτικής μεταγωγής και δρομολόγησης με τον δυναμικό έλεγχο των διαθέσιμων πόρων του δικτύου, με σκοπό να υπερκεράσει τις τρέχουσες προκλήσεις στα κέντρα δεδομένων. Για αυτό το σκοπό ακολουθείται η προσέγγιση της κάθετης ανάπτυξης (vertical development), ξεκινώντας από το επίπεδο δεδομένων (data plane) της αρχιτεκτονικής φτάνοντας μέχρι το επίπεδο ελέγχου (control plane) των πόρων του δικτύου, δομώντας έτσι μια πλήρως λειτουργική λύση δικτύωσης για τα Κέντρα Δεδομένων. Στο πλαίσιο αυτό, σχεδιάστηκαν και παρουσιάζονται δύο φιλόδοξες καινοτομίες:

- Το **επίπεδο δεδομένων (data plane)**, στο οποίο αξιοποιούνται εμπορικά διαθέσιμα φωτονικά στοιχεία σε συνδυασμό με τεχνικές διαίρεσης χρόνου (TDMA) με σκοπό την δυναμική και αποτελεσματική χρήση των πόρων του δικτύου [35, 36].
- Το **επίπεδο ελέγχου (control plane)** μαζί με ένα πλαίσιο ρύθμισης (orchestration framework) του δικτύου που βασίζεται σε πλατφόρμες Software-defined networking (SDN) και είναι υπεύθυνο για τη διαχείριση όλων των στοιχείων του επιπέδου δεδομένων. Στα πλαίσια της αρχιτεκτονικής, οι πλατφόρμες SDN που είναι ανοιχτού κώδικα, επεκτείνονται έτσι ώστε να υποστηρίζουν τις TDMA λειτουργίες στο δίκτυο. Από αυτή τη σκοπιά, το επίπεδο ελέγχου είναι ικανό να εκχωρεί δυναμικά πόρους δικτύου απευθείας στο οπτικό επίπεδο [37]. Επιπλέον αναπτύχθηκαν πολλές αλγοριθμικές προσθήκες που επικεντρώνονται στη γρήγορη κατανομή πόρων και που ενσωματώθηκαν στην SDN πλατφόρμα [38].

Η λειτουργία του πρωτοτύπου μίνι-Datacenter που βασίζεται στην προτεινόμενη υβριδική αρχιτεκτονική παρουσιάζεται στο ΚΕΦΑΛΑΙΟ 4. Τα δύο επίπεδα, δεδομένων και ελέγχου, συνεργάζονται και συνυπάρχουν απρόσκοπτα. Σε αυτή την ενότητα παρουσιάζονται οι βασικές καινοτομίες και οι λειτουργίες του δικτύου.

### 1.3.1 Επισκόπηση του Επιπέδου Δεδομένων (Data-plane)

Το δίκτυο που παρουσιάζεται υιοθετεί μια επίπεδη και κλιμακούμενη τοπολογία που χρησιμοποιεί ενεργά και παθητικά οπτικά στοιχεία για να ξεπεράσει τις αδυναμίες των ιεραρχικών τοπολογιών που χρησιμοποιούνται ευρέως στις ηλεκτρικές αρχιτεκτονικές των Δικτύων Κέντρων Δεδομένων (DCN). Το προτεινόμενο δίκτυο DCN αποτελείται από δύο στρώματα διακοπών: τους μεταγωγείς ToR (Top-of-Rack) και τους pod switches, όπως άλλωστε έχει ήδη αναφερθεί. Με αυτόν τον τρόπο, η τοπολογία εξυπηρετεί αποτελεσματικά τόσο την επικοινωνία στον κατακόρυφο άξονα όσο και στον οριζόντιο, κάτι που αποτελεί σημαντικό πλεονέκτημα σε σύγκριση με τις συμβατικές τοπολογίες DCN. Επιπλέον, ο αριθμός των δομικών στοιχείων του δικτύου αυξάνει γραμμικά ανάλογα με τον αριθμό των τερματικών.



*Εικόνα 5: Επισκόπηση του data plane της αρχιτεκτονικής δικτύου*

Στο επίπεδο δεδομένων, η διακίνησή τους βασίζεται σε τεχνικές δρομολόγησης διαίρεσης του χρόνου σε χρονοσχιμές-slots (Time-Division-Multiple-Access TDMA). Η διάρκεια των χρονοσχιμών είναι 200  $\mu\text{s}$  που ακολουθούνται από ένα διάστημα παύσης 10  $\mu\text{s}$ , το οποίο αξιοποιείται για τον συντονισμό και την επαναδιαμόρφωση όλων των πόρων του δικτύου. Ο χρόνος παύσης καθορίστηκε λαμβάνοντας υπόψη αφενός την απόκριση των μη ιδανικών ηλεκτρονικών στοιχείων (DC-coupled) και αφετέρου τον χρόνο που χρειάζεται ο

δέκτης του FPGA για να συγχρονίσει. Από την άλλη, η διάρκεια του κάθε slot καθορίστηκε στα 200  $\mu$ s προκειμένου να επιτυγχάνεται χρήση δικτύου στο 95%. Επιπλέον, όπως έχει ήδη αναφερθεί, η λειτουργία σε slots διευκολύνει τη δυναμική κατανομή πόρων του δικτύου. Η διαδικασία προγραμματισμού και άρα κατανομής των διαθέσιμων πόρων, γίνεται κατά χρονικές περιόδους, που ονομάζονται περίοδοι προγραμματισμού (scheduling periods). Έτσι λοιπόν ο χρόνος χωρίζεται σε περιόδους προγραμματισμού, με κάθε περίοδο να αποτελείται από 80 slots ( $200\mu\text{s} \times 80 = 16 \text{ ms}$ ). Στο τέλος κάθε περιόδου προγραμματισμού, ο «ενορχηστρωτής» του δικτύου υπολογίζει την επαναδιαμόρφωση του δικτύου για κάθε slot της επόμενης περιόδου προγραμματισμού. Η διαδικασία αυτή παρουσιάζεται σε επόμενη υποενότητα.

### 1.3.1.1 Pod switch

Ο pod switch της προτεινόμενης αρχιτεκτονικής είναι αυτός που διακρίνει την δρομολόγηση δεδομένων μεταξύ και εντός των pods (*inter-* και *intra-pod*) και βασίζεται σε δύο τύπους ενεργητικών οπτο-ηλεκτρονικών φωτονικών στοιχείων: σε έναν  $1 \times 2$  διακόπτη επιλογής μήκους κύματος (Wavelength Selective Switch - WSS) που βασίζεται στην προσέγγιση «αποπολυπλεξία, μεταγωγή και πολυπλεξία» και σε έναν γρήγορο  $1 \times 2$  οπτικό διακόπτη. Μια σχηματική εικόνα του pod switch δίνεται στην Εικόνα 5.

Στην κατεύθυνση από τους κεντρικούς υπολογιστές προς το οπτικό δίκτυο, οι γρήγοροι  $1 \times 2$  μεταγωγείς κατευθύνουν την κίνηση είτε σε άλλο ToR του ίδιου pod (*intra-pod*) είτε στον οπτικό δακτύλιο (*inter-pod*). Για χάρη ευκολίας αυτός ο γρήγορος διακόπτης ονομάζεται «διακόπτης καθορισμού *intra-* και *inter-* pod επικοινωνίας». Και στις δύο περιπτώσεις, *inter-* και *intra-pod* επικοινωνίας, χρησιμοποιούνται παθητικά οπτικά στοιχεία για τη δρομολόγηση των οπτικών ροών ανά μήκος κύματος στον κατάλληλο προορισμό, και πιο συγκεκριμένα  $W \times W$  δρομολογητές πλέγματος κυματοδηγών σε διάταξη (AWGR).

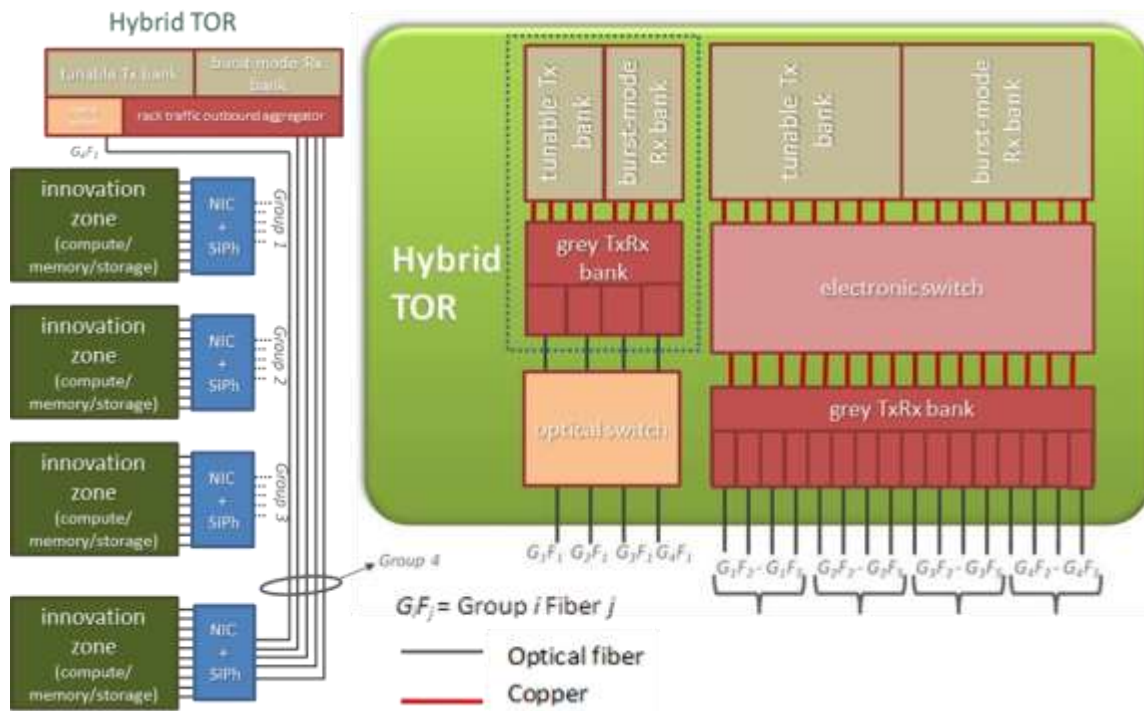
Στην κατεύθυνση από το οπτικό δίκτυο προς τους κεντρικούς υπολογιστές, ο  $1 \times 2$  διακόπτης επιλογής μήκους κύματος (WSS) «ρίχνει» ή προωθεί τα μήκη κύματος προερχόμενα από τους οπτικούς δακτυλίους στα pods, ανάλογα με τον προγραμματισμό και τον τελικό προορισμό. Τα μήκη κύματος δρομολογούνται στον ToR προορισμού και σε αυτήν την περίπτωση μέσω ενός δρομολογητή AWGR. Ο συνδυασμός αυτών των πολλαπλών φωτονικών στοιχείων παθητικού φιλτραρίσματος εντός του δικτύου, επιτρέπει την επαναχρησιμοποίηση των μηκών κύματος μεταξύ των pods, με αποτέλεσμα την επεκτασιμότητα του δικτύου πέρα από τον αριθμό που τυπικά επιτυγχάνεται στα DWDM συστήματα. Ο έλεγχος των ενεργών φωτονικών στοιχείων του pod switch γίνεται από τον SDN ελεγκτή του δικτύου (ενότητα 1.3.2) ενώ κατά τη διάρκεια των πειραμάτων που περιγράφονται στα ακόλουθα κεφάλαια, υλοποιείται σε FPGA πλατφόρμες.

### 1.3.1.2 Ο υβριδικός Top-of-Rack μεταγωγέας (Hybrid ToR switch)

Κάθε Top-of-Rack μεταγωγέας (ToR switch) διασυνδέει τους κεντρικούς υπολογιστές του κέντρου δεδομένων αφενός μεταξύ τους εντός των rack και αφετέρου με το ανώτερο επίπεδο δικτύου, που διαχειρίζονται οι pod switches.

Ένας από τους βασικούς στόχους της προτεινόμενης αρχιτεκτονικής ήταν να κάνει ένα βήμα μπροστά στην οπτική μεταγωγή, φέρνοντας την οπτική τεχνολογία μέσα στο ToR switch, έτσι ώστε να η κυκλοφορία εντός του rack να γίνεται με οπτικά μέσα, με τρόπο παρόμοιο με τις υπάρχουσες ηλεκτρικές επικοινωνίες μεταξύ των rack. Για το σκοπό αυτό, προτείνεται μια υβριδική αρχιτεκτονική στο ToR switch που επιτρέπει την πλήρη οπτική επικοινωνία μεταξύ των servers μέσα στο rack και βασίζεται σε ένα “smart-NIC” που μπορεί να υποστηρίξει οπτική λειτουργία. Αυτή η ιδέα της υβριδικής αρχιτεκτονικής απεικονίζεται στην Εικόνα 6 όπου φαίνονται τέσσερις ηλεκτρικές συνδέσεις (Electrical domain) και μια οπτική σύνδεση (Optical domain) για κάθε ζώνη καινοτομίας.

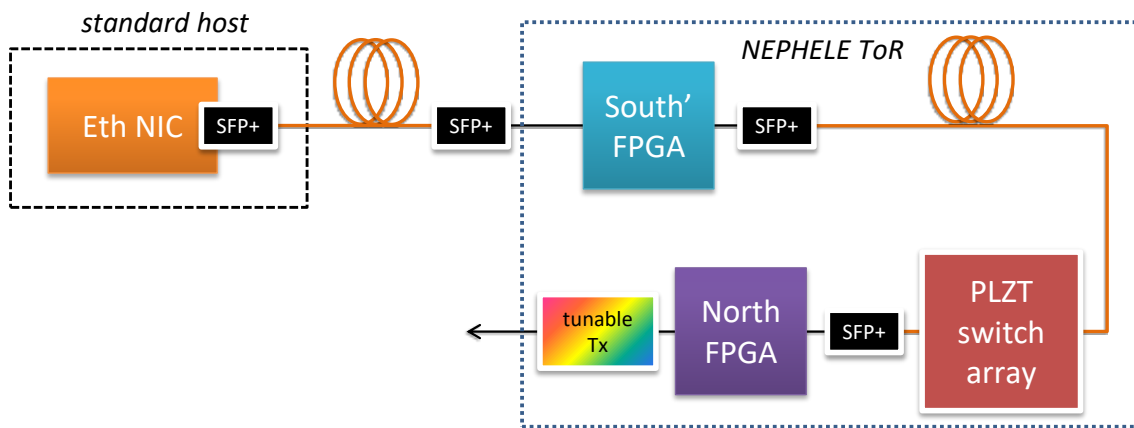
Συνολικά, σύμφωνα με την πρώτη ιδέα της αρχιτεκτονικής, ο ToR διακόπτης είναι διαστάσεων 20×20. Το ένα σκέλος του ToR switch υλοποιείται από έναν κλασικό Ethernet μεταγωγέα 16×16 συνδυασμένο με μια επέκταση που διαθέτει 16 θύρες προς τα pod-switches (south), καθώς και μια επέκταση που εκθέτει άλλες 16 θύρες Ethernet προς τους διακομιστές (north, NICs – Κάρτες Διασύνδεσης Δικτύου). Μια πρόσθετη μονάδα που διαθέτει 4 θύρες προς στα NIC και άλλες 4 θύρες για τους οπτικούς διακόπτες χρησιμοποιείται για την οπτική διασύνδεση των NIC απευθείας στους οπτικούς διακόπτες. Με αυτήν την υλοποίηση, έχουμε από τη μία 20 διεπαφές που συνδέονται με τα NIC με 10 τυπικές GbE Ethernet θύρες και από την άλλη 20 διεπαφές προς τα pod switches που είναι αποκλειστικά οπτικές θύρες που υποστηρίζουν την TDMA τεχνική.



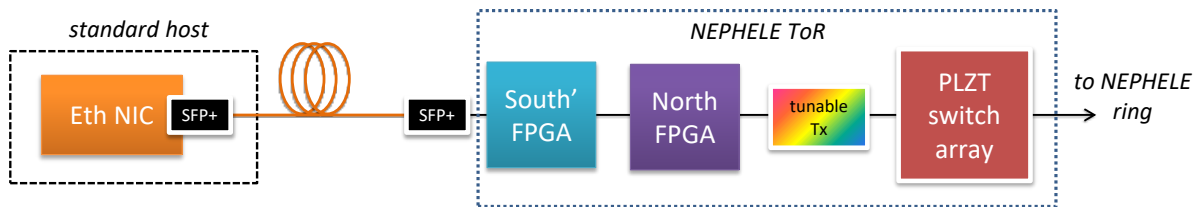
**Εικόνα 6: Η αρχική αρχιτεκτονική του Hybrid ToR switch**

Έπειτα από μελέτη που όμως δεν εμπίπτει στα πλαίσια της συγκεκριμένης διατριβής, η τελική εκδοχή της προτεινόμενης αρχιτεκτονικής, μεταφέρει τη λειτουργικότητα του smart-NIC στο υβριδικό ToR και πιο συγκεκριμένα στο κάτω μέρος του ToR (South interface) προκειμένου να υπάρχει συμβατότητα των υποδομών με τα τυπικά Ethernet NIC. Αυτό έχει σαν αποτέλεσμα πολλές από τις απαραίτητες λειτουργίες να γίνονται πλέον στο ToR switch, ενώ με αυτήν την επιλογή, οι θύρες ηλεκτρικού και οπτικού τομέα του ToR γίνονται και οι δύο συμβατές με τυπικούς κεντρικούς υπολογιστές Ethernet. Ως αποτέλεσμα, ένα τυπικό NIC Ethernet μπορεί να συνδεθεί απευθείας στο ToR switch.

Το γενικό διάγραμμα του οπτικού τμήματος του ToR switch φαίνεται στην Εικόνα 7. Ένας οπτικός διακόπτης  $N \times N$  εισάγεται μεταξύ δύο πλατφορμών FPGA (South FPGA και North FPGA), διευκολύνοντας την οπτική εναλλαγή εντός του ToR. Η απλοποιημένη εκδοχή της αρχιτεκτονικής φαίνεται στην Εικόνα 8. Ο οπτικός διακόπτης  $N \times N$  μεταφέρεται στη βόρεια διεπαφή του ToR και έτσι συνδέεται απευθείας με τους πομπούς του. Αυτή η προσέγγιση επιτρέπει την παράκαμψη των οπτικών συνδέσεων δεύτερου σταδίου της Εικόνα 7 (δηλαδή σύνδεση του South FPGA και του North FPGA) που απαιτούν SFP+ διεπαφές για να είναι συμβατά αφενός με τα σήματα TDMA και αφετέρου με την απλή λειτουργία του οπτικού διακόπτη  $N \times N$ . Στην ουσία, η λειτουργία αυτών των προσαρμοσμένων πομποδεκτών SFP+ συγχωνεύεται με την αντίστοιχη των πομποδεκτών στο North FPGA.



**Εικόνα 7: Γενικό διάγραμμα του οπτικού τμήματος του ToR switch**



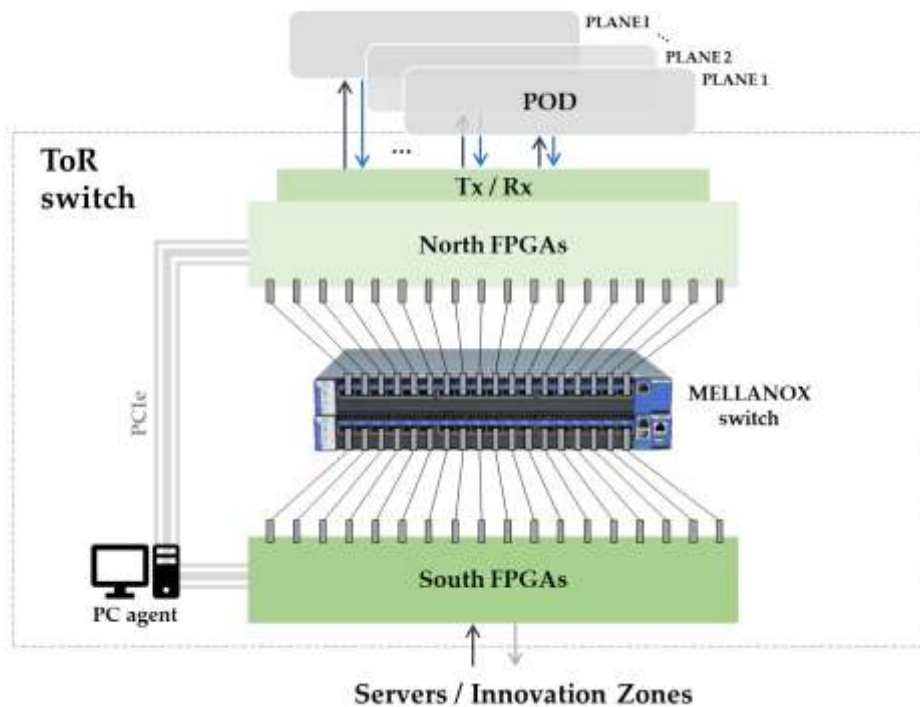
**Εικόνα 8: Απλοποιημένη εκδοχή του οπτικού τμήματος του ToR switch**

### 1.3.1.3 Απλοποιημένη υλοποίηση του ToR switch

Για να υποστηριχθεί η λειτουργία της οπτικής μεταγωγής σε slots, αναπτύχθηκαν ορισμένες επεκτάσεις στις βασικές λειτουργίες ενός εμπορικού ηλεκτρικού ToR διακόπτη. Οι επεκτάσεις αυτές υλοποιούνται σε FPGAs, ενώ όπως φαίνεται και στην Εικόνα 9 τοποθετούνται από τη μία ανάμεσα στον ηλεκτρικό διακόπτη και στους διακομιστές, για ευκολία το ονομάζουμε “South FGPA” και από την άλλη μεταξύ του ηλεκτρικού διακόπτη και του οπτικού δικτύου, “North FPGA”.

Με αυτόν τον τρόπο, σε ένα μελλοντικό Κέντρο Δεδομένων, οι ζώνες καινοτομίας και οι εφαρμογές που εκτελούνται σε αυτές θα παραμείνουν διαφανείς στους περιορισμούς του οπτικού δικτύου, ενώ ο ToR switch θα αναλάβει την απρόσκοπτη ενσωμάτωσή τους μέσω χρονοσχισμών στο οπτικό δίκτυο.





Εικόνα 9: Απλοποιημένη υλοποίηση του ηλεκτρικού τμήματος του ToR switch

### 1.3.1.3.1 Επέκταση South-FPGA (S-FPGA extender)

Το South FPGA βρίσκεται μεταξύ των διακομιστών και του ηλεκτρικού διακόπτη ToR. Λαμβάνει τα πλαίσια Ethernet που δημιουργούνται από τους διακομιστές στις ζώνες καινοτομίας, αναλύει τις κεφαλίδες τους και τις αποθηκεύει σε εικονικές ουρές VOQ (Virtual Output Queues) ανά ζεύγη ToR προορισμού και θύρας εισόδου. Για την καλύτερη κατανομή της διαθέσιμης μνήμης, τα VOQ δημιουργούνται δυναμικά ανάλογα με την εισερχόμενη κίνηση. Το South FPGA προωθεί μέρος των πλαισίων Ethernet προσθέτοντας μια ετικέτα εικονικού τοπικού δικτύου (VLAN Virtual Local Area Network) με τον ίδιο προορισμό ToR (δηλαδή από το ίδιο VOQ) στον Ethernet διακόπτη. Το South FPGA έχει ένα αμφίδρομο κανάλι επικοινωνίας με το επίπεδο ελέγχου. Πιο συγκεκριμένα, ειδοποιεί περιοδικά τον ελεγκτή SDN, και συνεπώς τον «ενορχηστρωτή» του δικτύου, σχετικά με την κατάσταση των VOQ. Στην αντίθετη κατεύθυνση, ο ελεγκτής SDN στέλνει τις ακόλουθες οδηγίες: α) ποιο VOQ θα αδειάσει για καθεμία από τις επερχόμενες χρονοσχισμές και β) την ετικέτα VLAN για κάθε χρονοσχισμή (η οποία μέσω του ηλεκτρικού διακόπτη ToR θα ορίσει την εξερχόμενη θύρα/επίπεδο για μετάδοση στο οπτικό δίκτυο).

### 1.3.1.3.2 Ηλεκτρικός Μεταγωγέας (Legacy electrical switch)

Στον ηλεκτρικό μεταγωγέα χρησιμοποιούνται δύο διαφορετικά σχήματα μεταγωγής, ανάλογα με την κατεύθυνση της κυκλοφορίας. Τα πλαίσια που λαμβάνονται από το South

FPGA, και επομένως με κατεύθυνση από τους διακομιστές προς το οπτικό δίκτυο (upstream), αλλάζουν με βάση την VLAN ετικέτα. Η VLAN ετικέτα έχει εισαχθεί στο South FPGA σύμφωνα με τις οδηγίες του ελεγκτή SDN για να κατευθύνει τα μέρη των πλαισίων ethernet στο κατάλληλο οπτικό επίπεδο. Στην αντίθετη κατεύθυνση, από το οπτικό δίκτυο προς τους διακομιστές (downstream), η εναλλαγή βασίζεται στη διεύθυνση προορισμού (MAC/IP). Με αυτόν τον τρόπο, τα πλαίσια που φτάνουν στον ToR switch από το οπτικό επίπεδο προωθούνται στους κατάλληλους servers ή ζώνες καινοτομίες.

### **1.3.1.3.3 Επέκταση North FPGA (N-FPGA extender)**

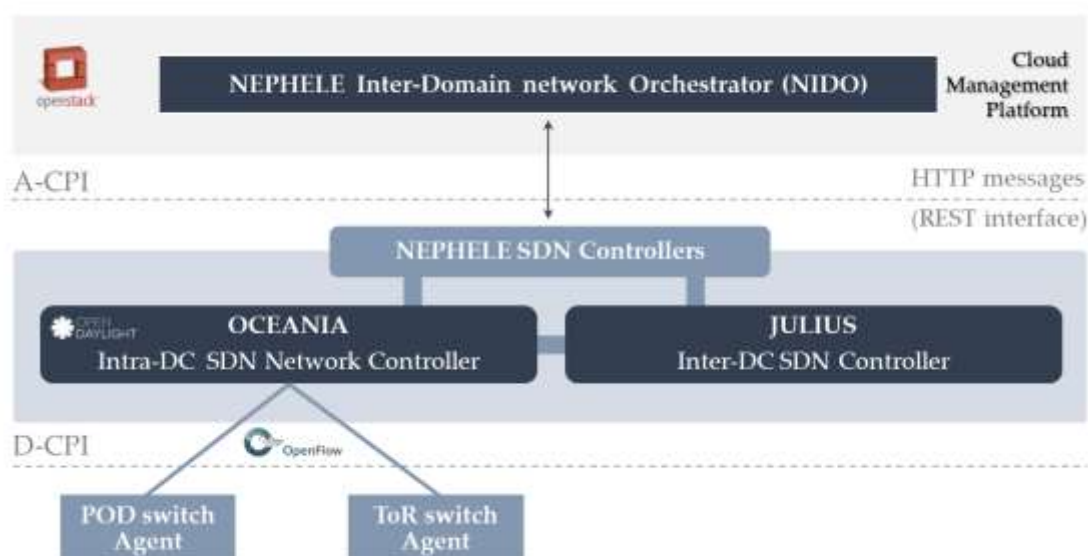
Όπως αναφέρθηκε και πιο πάνω, το North FPGA λαμβάνει τα Ethernet πλαίσια από τον ηλεκτρικό διακόπτη. Εν συνεχεία, στο N-FPGA γίνεται η παραμετροποίηση του οπτικού πομποδέκτη, ενώ παράλληλα υλοποιείται η λειτουργία σε χρονοσχισμές (slotted operation). Πιο συγκεκριμένα, η πλατφόρμα του N-FPGA extender περιλαμβάνει και μια μνήμη first-in-first-out (FIFO) στο μέγεθος μιας χρονοσχισμής (slot-sized) για την ακριβή ρύθμιση του χρονισμού των δεδομένων που μεταβιβάζονται στο οπτικό δίκτυο. Τα Ethernet πλαίσια ενσωματώνονται στο νέο πακέτο της προτεινόμενης αρχιτεκτονικής, στο οποίο προστίθεται ένα προοίμιο περίπου 8μs για την διευκόλυνση του «κλειδώματος» του ρολογιού ανάκτησης δεδομένων (Clock Data Recovery – CDR) στον δέκτη, και ορισμένες κεφαλίδες (headers) με πεδία διαχείρισης του δικτύου. Στην πλευρά της λήψης, τα νέο πακέτο «αποσυντίθεται» και τα Ethernet πλαίσια προωθούνται στον ηλεκτρικό διακόπτη.

Τέλος, το N-FPGA υλοποιεί την παραμετροποίηση του οπτικού πομποδέκτη, του οποίου βασικό δομικό στοιχείο είναι ο συντονιζόμενος σε μήκος κύματος πομπός (Tunable Tx), που αποτελείται από ένα συντονιζόμενο λείζερ, έναν διαμορφωτή MZI και έναν RF ενισχυτή (λεπτομέρειες στην ενότητα 2.1). Ο προσαρμοζόμενος δέκτης (Tx) είναι υπεύθυνος για την αποτύπωση των ηλεκτρικών δεδομένων των 10 Gb/s σε έναν οπτικό φορέα (μήκος κύματος). Ο πομπός στέλνει πακέτα ανά χρονοσχισμή σύμφωνα με την TDMA τεχνική. Το μήκος κύματος που εκπέμπεται σε κάθε χρονοσχισμή υπαγορεύεται στο N-FPGA από τον ελεγκτή SDN σύμφωνα με το ToR προορισμού.

### **1.3.2 Επισκόπηση του Επιπέδου Ελέγχου (Control Plane)**

Το πλαίσιο ελέγχου και εντοπισμού της προτεινόμενης αρχιτεκτονικής δεν εξετάζεται στα πλαίσια της παρούσας διατριβής, ωστόσο παρουσιάζουμε κάποια βασικά του στοιχεία. Πιο συγκεκριμένα, το πλαίσιο ελέγχου εφαρμόζεται τόσο στον τομέα εντός ενός DCN όσο και στον τομέα μεταξύ DC. Αυτό σημαίνει ότι το επίπεδο ελέγχου βασίζεται σε έναν κεντρικό εντοπιστή δικτύου μεταξύ των τομέων, που ονομάζεται NIDO [39] (NEPHELE Inter-Domain network Orchestrator) και ο οποίος συντονίζει πολλαπλά DCN και τους ενδιάμεσους κόμβους διασύνδεσης, με σκοπό την σωστή κατανομή πόρων από άκρο σε άκρο [40]. Πιο συγκεκριμένα, ο NIDO εντοπίζει τις ενέργειες των SDN

ελεγκτών εντός του τομέα κατώτερου επιπέδου, και συγκεκριμένα του OCEANIA [41] (OptiCal Electrical ApplicationN Aware data center controller) για την επικοινωνία εντός των DCN και του JULIUS [42] για την επικοινωνία μεταξύ DC, όπως απεικονίζεται στην Εικόνα 10. Αυτή η ιεραρχική προσέγγιση επιτρέπει την αλλαγή του ελεγκτή εντός του τομέα ή/και της σχετικής τεχνολογίας δικτύου οποιουδήποτε τομέα (DC, core, metro, access network).



**Εικόνα 10: Σχηματική απεικόνιση του επιπέδου ελέγχου της προτεινόμενης αρχιτεκτονικής**

Το εσωτερικό δίκτυο του Κέντρου Δεδομένων της προτεινόμενης αρχιτεκτονικής ελέγχεται μέσω του ελεγκτή OCEANiA, που στην ουσία είναι μια εξέλιξη του ελεγκτή ανοικτού κώδικα OpenDaylight (Lithium version) [43]. Ο OCEANiA συνδυάζει κι άλλες εφαρμογές SDN, αλγόριθμους και επεκτάσεις OpenFlow [44] πρωτοκόλλου για την αποτελεσματική λειτουργία ενός DCN, ενώ φροντίζει για τη δυναμική εγκαθίδρυση, επαναδιαμόρφωση ή διακοπή όλων των συνδέσεων εντός DC βελτιστοποιώντας την κατανομή των πόρων στο χώρο (optical planes) και στο χρόνο (timeslots).

Κατά την ανάπτυξη του πρωτοτύπου μίνι-DCN (Εικόνα 11) που παρουσιάζεται στα πλαίσια αυτής της διατριβής, επιτεύχθηκε η ενσωμάτωση του ελεγκτή OCEANiA με το επίπεδο δεδομένων στο south-bound και του «ενορχηστρωτή» NIDO στο north-bound. Όσον αφορά στις SDN εφαρμογές, υλοποιήθηκαν αλγόριθμοι κατανομής πόρων (scheduling) για να καταστεί δυνατή η αποτελεσματική και πλήρως αυτοματοποιημένη δημιουργία οπτικών συνδέσεων. Επιπλέον έγινε προσπάθεια ανάπτυξης αποτελεσματικών αλγορίθμων χρονοπρογραμματισμού για μεγάλα DCN που επιτρέπουν γρήγορους υπολογισμούς για την ταχεία αναδιαμόρφωση του δικτύου [38] [45].



*Εικόνα 11: Φωτογραφίες του μίνι datacenter κατά την περίοδο της προετοιμασίας. (α) Ο διακόπτης rod, συμπεριλαμβανομένου ενός WSS. (β) Τα γρήγορα λέιζερ που χρησιμοποιούνται για τον πομπό. (γ) Τα FPGA extenders του ToR διακόπτη*

## ΚΕΦΑΛΑΙΟ 2

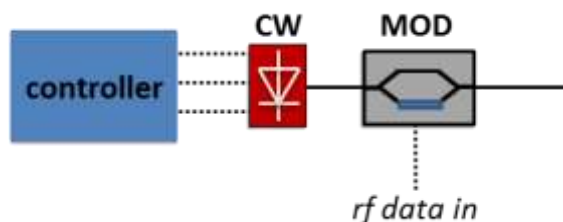
# Σχεδιασμός Οπτικών Υποσυστημάτων

Όπως αναφέρθηκε και στο ΚΕΦΑΛΑΙΟ 1, που περιγράφει τις αρχές που διέπουν και τα μέρη που απαρτίζουν την προτεινόμενη αρχιτεκτονική, ο Top-of-Rack διακόπτης (εν συντομία *ToR switch*) αποτελεί βασικό σύστημα. Ο ToR switch διασυνδέει τους κεντρικούς υπολογιστές που βρίσκονται στα racks του κέντρου δεδομένων με το ανώτερο επίπεδο δικτύου, το οποίο διαχειρίζονται οι rod switches. Σε αυτό το κεφάλαιο περιγράφονται τα επιμέρους οπτικά υποσυστήματα του ToR switch ενώ για λόγους πληρότητας γίνεται αναφορά και σε αυτά που απαρτίζουν τον rod switch, παρόλο που δεν είναι αντικείμενο της παρούσας διδακτορικής διατριβής.

Τα μέρη που αποτελούν τον ToR switch και αναλύονται στο παρόν κεφάλαιο είναι ο προσαρμοζόμενος πομπός, ο δέκτης και ο γρήγορος οπτικός διακόπτης για εξολοκλήρου οπτική μεταγωγή. Επίσης, αναφορικά με τον rod switch παρουσιάζονται ο διακόπτης επιλογής δρομολόγησης, τα παθητικά φίλτρα δρομολόγησης και ο διακόπτης επιλογής μήκους κύματος (WSS). Ο επιτυχής σχεδιασμός των οπτικών υποσυστημάτων αξιολογείται μέσω πειραματικών διατάξεων που παρουσιάζονται σε αυτό το κεφάλαιο.

### 2.1 Προσαρμοζόμενος Πομπός στο ToR switch

Ο δυναμικά προσαρμοζόμενος πομπός βρίσκεται στο μεταγωγέα ToR switch της προτεινόμενης υβριδικής αρχιτεκτονικής και χρησιμεύει για την αποτύπωση σε έναν οπτικό φορέα (optical carrier) των ψηφιακών δεδομένων που προέρχονται από τα τερματικά (endhosts) που είναι συνδεδεμένα με το ToR. Το μήκος κύματος του οπτικού φορέα φέρει τις πληροφορίες δρομολόγησης για τη συγκεκριμένη ροή και επομένως επιλέγεται μεταξύ ενός προκαθορισμένου πλέγματος μηκών κύματος.



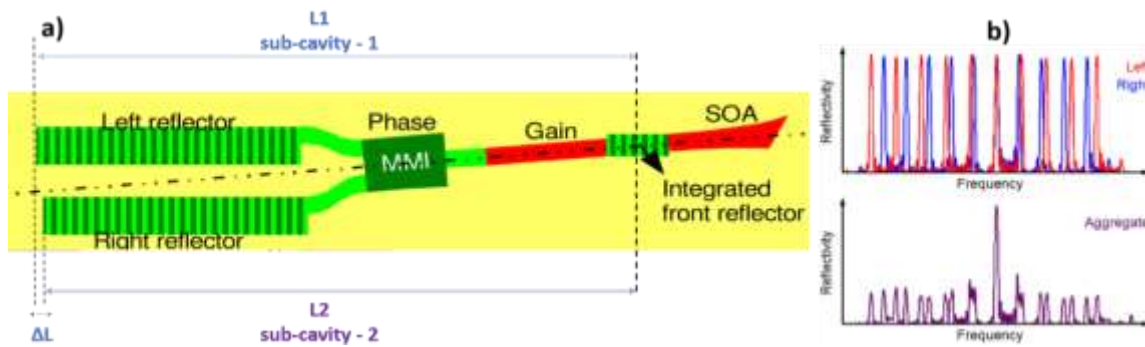
*Εικόνα 12: Διάγραμμα του πομπού της προτεινόμενης υβριδικής αρχιτεκτονικής.*

Το διάγραμμα του πομπού απεικονίζεται στην Εικόνα 12, όπου διακρίνονται τα δύο οπτοηλεκτρονικά δομικά στοιχεία, δηλαδή η οπτική πηγή (laser) και ο οπτικός διαμορφωτής (Modulator), καθώς και τα κυκλώματα ελέγχου για την επιθυμητή ρύθμιση του laser. Το laser που χρησιμοποιείται είναι προσαρμοζόμενο (tunable), δηλαδή μπορεί να παράξει συνεχές οπτικό κύμα (CW) στο επιθυμητό μήκος κύματος. Εν συνεχεία το φως εισάγεται στον οπτικό διαμορφωτή όπου τα δεδομένα που προέρχονται από τους endhosts εγγράφονται στον οπτικό φορέα. Ο σχεδιασμός και ο χαρακτηρισμός του προσαρμοζόμενου πομπού έγινε σταδιακά, όπως περιγράφεται λεπτομερώς στις ακόλουθες παραγράφους.

### 2.1.1 Επισκόπηση και αρχή λειτουργίας της προσαρμοζόμενης οπτικής πηγής λέιζερ (Tunable Laser)

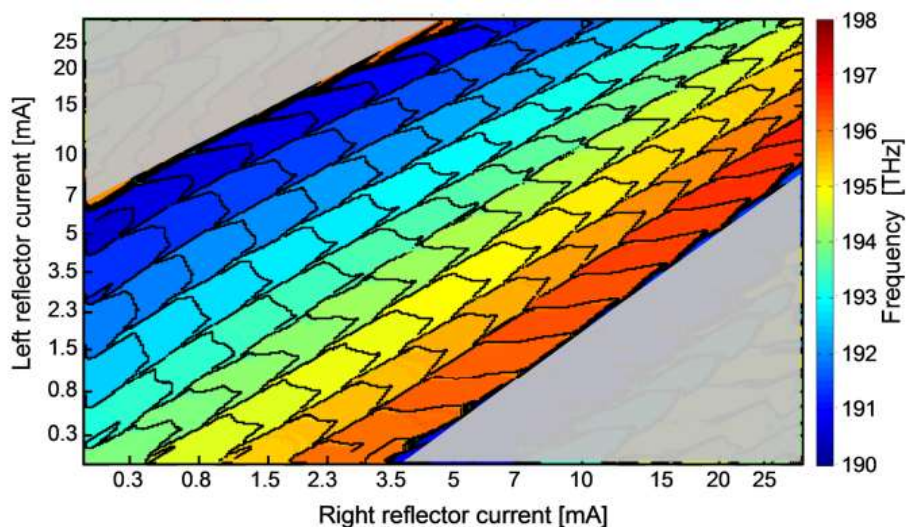
Στα πλαίσια της μελέτης χρησιμοποιήθηκε το προσαρμοζόμενο λέιζερ S7500 της εταιρίας Finisar που είναι εμπορικά διαθέσιμο και η λειτουργία του είναι βασισμένη σε ένα διαμορφωμένο πλέγμα της Y-διακλάδωσης (Modulated Grating Y-branch structure, MG-Y) [46]. Αυτός ο σχεδιασμός αποτελεί μια παραλλαγή του συμβατικού κατανεμημένου λέιζερ Bragg reflector (DBR) [47] [48]. Παρόλα αυτά, όπως φαίνεται και στην Εικόνα 13 (α) η ενσωμάτωση ενός επιπρόσθετου τμήματος υπο-κοιλότητας διασφαλίζει ένα εκτεταμένο εύρος συντονισμού μήκους κύματος σε ολόκληρη τη C-band, δίνοντας έναν επιπλέον βαθμό ελευθερίας για τη διαδικασία συντονισμού μήκους κύματος σε αντίθεση με τα τυπικά λέιζερ DBR που περιορίζονται σε  $\sim 1\text{ THz}$  ( $\sim 8\text{ nm}$ ).

Ως εκ τούτου, ο συντονισμός μήκους κύματος του λέιζερ MG-Y βασίζεται στο φαινόμενο Vernier [48]. Οι δύο διαμορφωμένοι ανακλαστήρες σχάρας έχουν φάσμα ανακλαστικότητας σε σχήμα χτένας, όπως απεικονίζεται στο άνω μισό τμήμα του σχήματος στην Εικόνα 13 (β). Οι χτένες έχουν ελαφρώς διαφορετικές κορυφές, έτσι ώστε μόνο ένα ζευγάρι κορυφών να επικαλύπτεται ανά πάσα στιγμή. Και οι δύο ανακλάσεις συνδυάζονται χρησιμοποιώντας ένα ζεύκτη συμβολής πολλαπλών τρόπων (multi-mode interference - MMI). Η συνολική αντανάκλαση που φαίνεται από τη θύρα εισόδου του συζεύκτη MMI απεικονίζεται στο κάτω μισό τμήμα του σχήματος στην Εικόνα 13 (β). Μια μεγάλη ανάκλαση εμφανίζεται μόνο στη συχνότητα όπου μια κορυφή ανακλαστικότητας από τον αριστερό ανακλαστήρα συμπίπτει με μια κορυφή ανακλαστικότητας από τον δεξιό ανακλαστήρα. Το λέιζερ θα εκπέμπει έτσι φως στη συχνότητα της διαμήκης λειτουργίας κοιλότητας που είναι πλησιέστερα στην κορυφή της συνολικής αντανάκλασης. Ρυθμίζοντας έναν από τους ανακλαστήρες σε διάστημα ίσο με τη διαφορά στο διαχωρισμό κορυφής, ένα γειτονικό ζεύγος κορυφών μπορεί να ευθυγραμμιστεί, δηλαδή ένας μεγάλος συντονισμός της συχνότητας εκπομπής (μήκος κύματος) επιτυγχάνεται για έναν σχετικά μικρό συντονισμό ενός μόνο ανακλαστήρα. Πρέπει να σημειωθεί ότι το τμήμα *Gain* (κέρδος) του λέιζερ είναι μέρος της κοιλότητας του λέιζερ, ενώ το τμήμα *SOA* χρησιμοποιείται εξωτερικά για την ενίσχυση της επιλεγμένης λειτουργίας μήκους κύματος.



**Εικόνα 13:** α) Σχηματική διάταξη της δομής λέιζερ MG-Y που χρησιμοποιείται στην προτινόμενη υβριδική αρχιτεκτονική μεταγωγής, β) (πάνω) δύο διαφορετικές χτένες συχνότητας με ελαφρώς διαφορετικό εύρος ελεύθερου φάσματος (FSR) που προέρχεται από τα διαφορετικά μήκη υπο-κοιλότητας, (κάτω) συνολική απόκριση συχνότητας του λέιζερ με βάση το φαινόμενο Vernier.

Ένας ακριβής χαρακτηρισμός του μήκους κύματος ως συνάρτηση των ρευμάτων συντονισμού παρέχεται από τον κατασκευαστή και φαίνεται στην Εικόνα 14, όπου απεικονίζεται το μήκος κύματος εξόδου έναντι του συνδυασμού των ρευμάτων αριστερού και δεξιού ανακλαστήρα, για μια σταθερή τιμή ρεύματος στο τμήμα της φάσης (*phase section*). Είναι σαφές ότι προσδιορίζονται μεγάλες περιοχές (super-modes / υπερ-ρυθμοί) που εμφανίζουν απότομες αλλαγές μήκους κύματος μεταξύ τους. Εντός των super-modes, εντοπίζονται οι πιο σταθερές περιοχές, δείχνοντας ότι κάθε super-mode διαιρείται σε διάφορες περιοχές τραπεζοειδούς σχήματος (ρυθμούς / modes), οι οποίες αντιπροσωπεύουν τις ζώνες σταθερού μήκους κύματος, με λόγο συμπίεσης πλευρικής λειτουργίας (side-mode suppression ratio SMSR) > 40 dB.

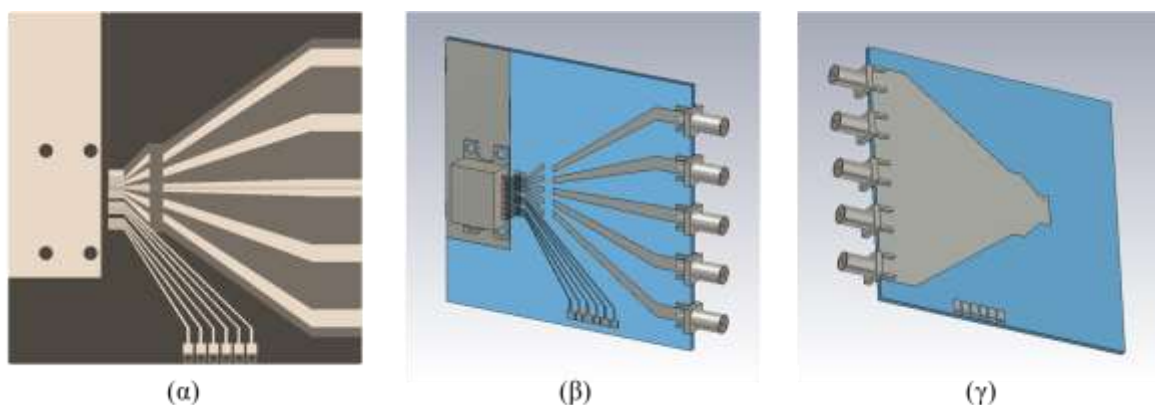


**Εικόνα 14:** Συχνότητα εκπομπής λέιζερ MG-Y που απεικονίζεται σε συνάρτηση με τα αριστερά και δεξιά ρεύματα ανακλαστήρα. Μια ασυνεχής αλλαγή συχνότητας συμβαίνει στα όρια που υποδεικνύονται από τα μαύρα περιγράμματα [46]

Οι πληροφορίες σχετικά με αυτές τις ζώνες επιτρέπουν τον προσδιορισμό των περιοχών λειτουργίας λέιζερ που αντιστοιχούν στα κανάλια που καθορίζονται από το ITU-T. Πρέπει να σημειωθεί ότι με την έγχυση ρεύματος στο τμήμα φάσης, η φάση της κοιλότητας αλλάζει. Έτσι, αυξάνοντας το ρεύμα φάσης, όλες οι περιοχές λειτουργίας στο χάρτη που φαίνονται στην Εικόνα 14 μετατοπίζονται προς υψηλότερα ρεύματα ανακλαστήρα.

### 2.1.2 Ρύθμιση λέιζερ και ενσωμάτωση σε πλακέτα (PCB)

Το λέιζερ S7500 της Finisar διατίθεται σε μια συμπαγή μορφή (compact butterfly package) και συνοδεύεται από μια ξεχωριστή ηλεκτρονική πλακέτα που συνδέεται με υπολογιστή για τον έλεγχο του μήκους κύματος εξόδου. Παρόλο που η εγγενής ταχύτητα συντονισμού του μήκους κύματος του λέιζερ MG-Y βρίσκεται στην κλίμακα νανοδευτερολέπτων, η ηλεκτρονική πλατφόρμα που το συνοδεύει είναι ικανή να συντονίζει το μήκος κύματος του λέιζερ μέσα σε αρκετά χιλιοστά του δευτερολέπτου, κάτι που δεν είναι κατάλληλο για τις ανάγκες της προτεινόμενης υβριδικής αρχιτεκτονικής. Επομένως, κρίθηκε απαραίτητη η ανάπτυξη ενός εξατομικευμένου ηλεκτρονικού κυκλώματος οδήγησης του λέιζερ, προκειμένου να αξιοποιηθούν πλήρως τα δυναμικά χαρακτηριστικά του.

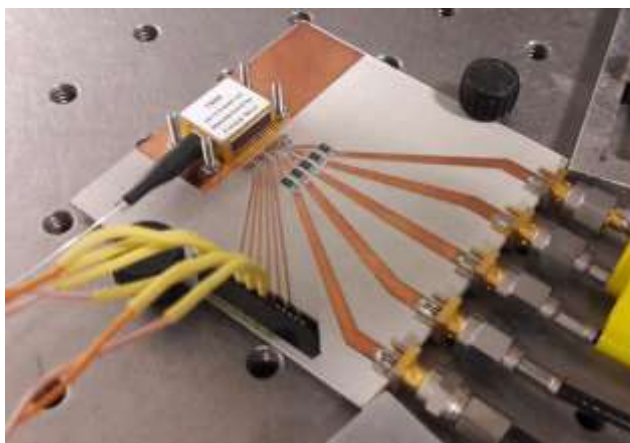


**Εικόνα 15:** (α) Το σχέδιο δύο διαστάσεων του κυκλώματος της πλακέτας (PCB), (β-γ) Απεικόνιση τριών διαστάσεων (3D) της πλακέτας με την χρήση σχεδιαστικού προγράμματος.

Για την σωστή ενσωμάτωση του λέιζερ σχεδιάστηκε μια τυπωμένη ηλεκτρονική πλακέτα κυκλώματος (Printed circuit board - PCB) χρησιμοποιώντας λογισμικό ηλεκτρονικής σχεδίασης (Εικόνα 15). Το PCB παρέχει πέντε γραμμές RF υψηλής ταχύτητας για τον ανεξάρτητο έλεγχο των πέντε προσαρμοζόμενων τμημάτων του λέιζερ: *gain*, *SOA*, *phase*

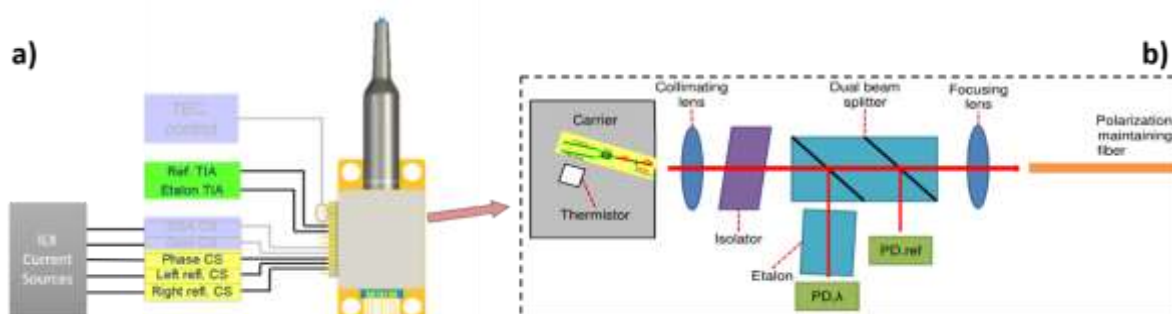


και των δύο *reflectors* (ανακλαστήρων), καθώς και έξι γραμμές συνεχούς ρεύματος (direct current - DC) για των έλεγχο των διασυνδέσεων χαμηλότερης ταχύτητας του λέιζερ, δηλαδή των εξής τμημάτων: Ελεγκτής θερμοκρασίας (Temperature controller – TEC), θερμίστορ (Thermistor) και τμήματα φωτοδιόδου παρακολούθησης (monitor photodiode). Για να εξασφαλιστεί γρήγορος συντονισμός μήκους κύματος, οι γραμμές υψηλής ταχύτητας στο PCB σχεδιάστηκαν για να ταιριάζουν την αντίσταση εισόδου του λέιζερ με την αντίσταση εξόδου των ηλεκτρονικών οδήγησης (50 Ohm). Μετά την κατασκευή PCB, το λέιζερ τοποθετήθηκε αναλόγως και συγκολλήθηκε πάνω στην πλακέτα όπως φαίνεται στην Εικόνα 16.



**Εικόνα 16:** Προσαρμοζόμενο λέιζερ στο PCB. Στη φωτογραφία είναι ορατές οι γραμμές RF υψηλής ταχύτητας αλλά και οι γραμμές χαμηλής ταχύτητας που τερματίζουν σε SMA υποδοχείς.

Το σχεδιάγραμμα με τις συνδέσεις του προσαρμοζόμενου λέιζερ φαίνεται στην Εικόνα 17 (α) όπως αυτό περιγράφεται από τον κατασκευαστή ενώ στην Εικόνα 17 (β) περιγράφονται τα εσωτερικά τμήματα του λέιζερ που σχετίζονται με την κυματοδότηση του φωτός εσωτερικά της συσκευής.



**Εικόνα 17:** Σχεδιαγράμματα α) Συνδέσεων των ρευμάτων στο Finisar S7500 λέιζερ και β) Εσωτερικών δομών της συσκευής.

Όπως συμβαίνει σε όλα τα λέιζερ ημιαγωγών, η σταθερότητα θερμοκρασίας παίζει καθοριστικό ρόλο στην ρύθμιση του μήκους κύματος του εκπεμπόμενου φωτός. Για τη σταθεροποίηση, λοιπόν, του μήκους κύματος, χρειάζεται ένας ελεγκτής θερμοκρασίας (temperature controller - TEC) καθώς και ένα θερμίστορ (thermistor - TH) για την παρατήρηση της θερμοκρασίας. Για το λόγο αυτό, ένα από τα πρώτα βήματα της διαδικασίας αξιολόγησης του λέιζερ αφορούσε τον υπολογισμό των συντελεστών θερμίστορ (C1, C2, C3) σύμφωνα με την εξίσωση Steinhart-Hart που δίνεται παρακάτω:

$$\frac{1}{T} = A + B \cdot \ln(R) + (C \cdot \ln(R))^3,$$

όπου το R υποδηλώνει την αντίσταση του θερμίστορ (Ohms) σε μια δεδομένη θερμοκρασία T (Kelvin) και τα A, B, C αντιστοιχούν στις σταθερές του θερμίστορ. Όσον αφορά στο προσαρμοζόμενο λέιζερ Finisar S7500, οι συντελεστές θερμίστορ που καθορίζονται στο φύλλο δεδομένων μπορούν να υπολογιστούν αναλυτικά από:

$$R = R_0 \cdot \exp\left(B \cdot \left(\frac{1}{T} - \frac{1}{T_0}\right)\right),$$

όπου:  $R_0 = 10 \text{ kOhm}$ ,  $B = 3930 \text{ K}$ ,  $T_0 = 298.15 \text{ K}$ , σύμφωνα με τον κατασκευαστή.

Χρησιμοποιώντας την παραπάνω εξίσωση προσδιορίστηκαν οι τιμές αντίστασης για συγκεκριμένες θερμοκρασίες, όπως φαίνεται στον παρακάτω πίνακα (Πίνακας 2).

**Πίνακας 2:** Υπολογισμένες τιμές της αντίστασης του θερμίστορ του λέιζερ για διαφορετικές τιμές θερμοκρασίας

T(°C)	T(°K)	R (Ohms)
<b>0</b>	273.15	33414.82
<b>30</b>	303.15	8046.04
<b>45</b>	318.15	4366.51

Τέλος, επιλύοντας την εξίσωση Steinhart-Hart [49] υπολογίζονται οι συντελεστές θερμίστορ C1, C2 και C3:

$$C1 = 1.01042705822087$$

$$C2 = 2.54451540180550$$

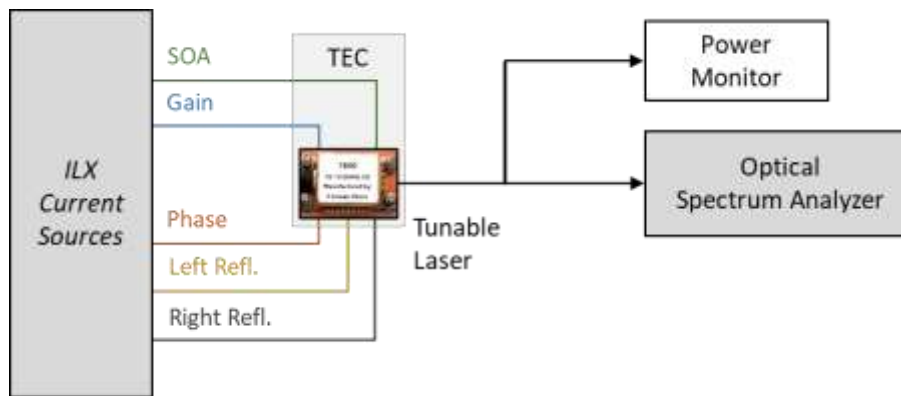
$$C3 = 0.00005104665711$$

Οι συντελεστές αυτοί χρησιμοποιήθηκαν σε έναν ελεγκτή θερμοκρασίας λέιζερ που συνδέθηκε με τους ακροδέκτες TEC και θερμίστορ του λέιζερ μέσω της πλακέτας. Ο ελεγκτής ρυθμίστηκε έτσι ώστε να διατηρεί τη θερμοκρασία του λέιζερ στους 25°C, η οποία

διατηρήθηκε σταθερή καθόλη τη διάρκεια των πειραμάτων που αναφέρονται στις επόμενες ενότητες.

### 2.1.3 Στατικός χαρακτηρισμός του συντονισμού μήκους κύματος του προσαρμοζόμενου λέιζερ στη C-band

Το πρώτο βήμα για τον χαρακτηρισμό του προσαρμοζόμενου λέιζερ αφορά στην εκτίμηση των συνθηκών οδήγησης για το συντονισμό του λέιζερ σε κάθε ένα από τα 80 κανάλια της C-band σύμφωνα με τα κεντρικά μήκη κύματος που καθορίζονται από το ITU-T [50]. Η Εικόνα 18 απεικονίζει την πειραματική διάταξη που χρησιμοποιήθηκε για τον χαρακτηρισμό του συντονισμού του μήκους κύματος του λέιζερ S7500 σε συνεχή λειτουργία (continuous wave - CW) στη φασματική περιοχή της C μπάντας, και πιο συγκεκριμένα στην περιοχή μεταξύ 1529,55nm - 1561,1nm.



**Εικόνα 18:** Πειραματική διάταξη που εφαρμόστηκε για τη ρύθμιση συντονισμού του μήκους κύματος και τη σταθερότητα ισχύος του λέιζερ S7500.

Η πηγή ρεύματος που χρησιμοποιήθηκε για την οδήγηση του προσαρμοζόμενου λέιζερ μέσω του κυκλώματος οδήγησης PCB (που περιεγράφηκε στην προηγούμενη παράγραφο) ήταν ένας ελεγκτής διόδων λέιζερ υψηλής ακρίβειας 16 καναλιών της ILX Lightwave (LDC-3916) που ελέγχει όλες τις γραμμές DC και RF ανεξάρτητα αλλά και ταυτόχρονα. Στην έξοδο του λέιζερ χρησιμοποιήθηκε ένας μετρητής οπτικής ισχύος (optical power meter) και ένας αναλυτής οπτικού φάσματος (optical spectrum analyzer – OSA) υψηλής ακρίβειας με ανάλυση 10pm, προκειμένου να μετρηθεί η ισχύς και να καταγραφεί το σχήμα φάσματος καθενός από τα εκπεμπόμενα μήκη κύματος.

Η ρύθμιση του συγκεκριμένου λέιζερ έχει πέντε διαφορετικούς βαθμούς ελευθερίας που επηρεάζουν τον συντονισμό του μήκους κύματος, και όπως αναφέρθηκε και προηγουμένως αυτοί είναι οι εξής: οι ανακλαστήρες δεξιά και αριστερά, το τμήμα φάσης, το τμήμα κέρδους και η θερμοκρασία του λέιζερ. Για να μειωθεί η πολυπλοκότητα της πειραματικής

διαδικασίας αξιολόγησης χωρίς να τροποποιηθούν ταυτόχρονα όλες οι πιθανές παράμετροι, καθορίστηκαν τα ρεύματα ελέγχου για το τμήμα του κέρδους (Gain) και το SOA στα 40 mA αντίστοιχα, ενώ η θερμοκρασία σταθεροποιήθηκε μέσω του TEC στους 25°C.

Σε αυτό το σημείο πρέπει να σημειωθεί ότι διαφορετικοί συνδυασμοί των προαναφερθεισών παραμέτρων θα μπορούσαν να οδηγήσουν στο ίδιο μήκος κύματος που εκπέμπεται από το λέιζερ. Αυτό μπορεί εύκολα να εξηγηθεί λόγω της περιοδικής φασματικής απόκρισης τόσο των ανακλαστήρων του λέιζερ όσο και της εγγενούς ιδιότητας του φαινομένου Vernier [51]. Αυτό παρατηρήθηκε επίσης στην πειραματική διαδικασία όπου τρεις από τις πέντε παραμέτρους ρύθμισης του λέιζερ είχαν καθοριστεί.

Επιπλέον, υπάρχει μια ορισμένη ανοχή στο εύρος των ρευμάτων συντονισμού για τους ανακλαστήρες και το τμήμα της φάσης του λέιζερ. Με άλλα λόγια, τα ρεύματα για ένα συγκεκριμένο μήκος κύματος μπορούν να αποκλίνουν γύρω από μια κεντρική τιμή και παρόλα αυτά το λέιζερ εξακολουθεί να μπορεί να εκπέμπει με υψηλή αναλογία συμπίεσης πλευρικής λειτουργίας (side mode suppression ratio SMSR). Το SMSR είναι υψηλότερο από 40 dB στο μεγαλύτερο μέρος αυτής της περιοχής ενώ μειώνεται μόνο κοντά στα όρια. Αυτή η συμπεριφορά περιγράφεται και από τον κατασκευαστή [46] και δείχθηκε επίσης πειραματικά κατά τη διαδικασία χαρακτηρισμού του λέιζερ στα πλαίσια των πειραμάτων της διατριβής. Αν και η λειτουργία του λέιζερ δεν επηρεάζεται από διακυμάνσεις στα ρεύματα οδήγησης, το μήκος κύματος εξόδου μπορεί να μεταβάλλεται κατά αρκετά GHz, κάτι που είναι ανεπιθύμητο δεδομένου ότι το δίκτυο της προτεινόμενης υβριδικής αρχιτεκτονικής χρησιμοποιεί ζωνοπερατά οπτικά φίλτρα που είναι αρκετά στενά (50GHz-100GHz). Για αυτόν τον λόγο, οι ακριβείς ρυθμίσεις που καταγράφηκαν για το συντονισμό του λέιζερ έγιναν ακριβώς στο ονομαστικό μήκος κύματος όπως καθορίζεται από το ITU-T [50].

Ο Πίνακας 3 δείχνει τα 80 κανάλια της C μάντας που αφορούν σε πολυπλεξία διαίρεσης μήκους κύματος (Dense wavelength division multiplexing – DWDM) σύμφωνα με την τυποποίηση της ITU-T για κανάλια με απόσταση 50 GHz (~ 0,4 nm) ενώ εμφανίζεται και η αντιστοιχία των καναλιών σε nm και σε THz. Οι στήλες «Phase», «Left» και «Right» αντιστοιχούν στις τιμές ρεύματος έγχυσης που εφαρμόζονται στη φάση και τους ανακλαστήρες του λέιζερ S7500 μέσω της ειδικά σχεδιασμένης πλακέτας RF/DC. Πρέπει να σημειωθεί ότι για όλα τα κανάλια DWDM που δίδονται στον παρακάτω πίνακα, οι αντίστοιχες τιμές ρεύματος για τους ανακλαστήρες επαληθεύτηκαν ότι είναι επαναλαμβανόμενες οδηγώντας σε περίπου το ίδιο μήκος κύματος ή συχνότητα με τουλάχιστον 8 dBm οπτική ισχύ και SMSR > 40 dB για κάθε μέτρηση.

***Πίνακας 3:** Πίνακας συντονισμού λέιζερ, που δείχνει τα αντίστοιχα ρεύματα για: αριστερό και δεξιό ανακλαστήρα και τμήμα φάσης, αντίστοιχα. Η θερμοκρασία μέτρησης ήταν στους 25°C και το ρεύμα και για τα τμήματα κέρδους και SOA ορίστηκε στα 40 mA.*

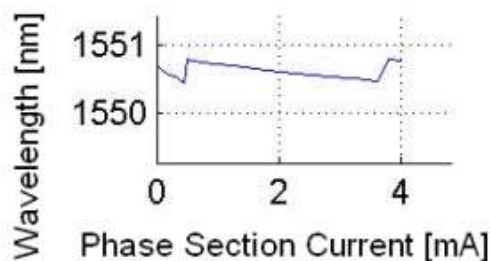
<b>Κανάλι</b>	<b>Συχνότητα (THz)</b>	<b>Μήκος Κόματος (nm)</b>	<b>Ρεύμα Φάσης (mA)</b>	<b>Ρεύμα αριστερού αν/ρα (mA)</b>	<b>Ρεύμα δεξιού αν/ρα (mA)</b>
<b>1</b>	192,050	1561,01	1,96	2,68	2,67
<b>2</b>	192,100	1560,61	4,52	3,26	3,25
<b>3</b>	192,150	1560,20	3,00	4,13	4,04
<b>4</b>	192,200	1559,79	1,71	5,13	4,96
<b>5</b>	192,250	1559,39	3,51	6,42	6,13
<b>6</b>	192,300	1558,98	4,85	7,96	7,35
<b>7</b>	192,350	1558,58	1,24	9,21	8,66
<b>8</b>	192,400	1558,17	0,94	10,39	9,99
<b>9</b>	192,450	1557,77	2,53	12,23	11,7
<b>10</b>	192,500	1557,36	1,52	14,17	13,72
<b>11</b>	192,550	1556,96	4,31	15,67	15,26
<b>12</b>	192,600	1556,55	1,86	19,50	19,09
<b>13</b>	192,650	1556,15	1,01	22,52	21,98
<b>14</b>	192,700	1555,75	2,53	25,30	24,84
<b>15</b>	192,750	1555,34	0,85	30,14	29,56
<b>16</b>	192,800	1554,94	1,34	3,26	4,18
<b>17</b>	192,850	1554,54	1,50	3,49	4,76
<b>18</b>	192,900	1554,13	3,92	4,37	5,75
<b>19</b>	192,950	1553,73	2,65	5,19	6,78
<b>20</b>	193,000	1553,33	5,49	6,37	8,11
<b>21</b>	193,050	1552,93	0,69	7,89	9,68
<b>22</b>	193,100	1552,52	2,42	8,96	11,22
<b>23</b>	193,150	1552,12	5,61	10,59	13,01
<b>24</b>	193,200	1551,72	3,67	12,19	15,02
<b>25</b>	193,250	1551,32	1,98	14,7	17,86
<b>26</b>	193,300	1550,92	4,1	17,43	20,59
<b>27</b>	193,350	1550,52	2,3	20,4	23,47
<b>28</b>	193,400	1550,12	5,1	22,27	25,69
<b>29</b>	193,450	1549,72	4,5	2,15	4,08
<b>30</b>	193,500	1549,32	0,83	2,45	4,66
<b>31</b>	193,550	1548,91	2,49	3,1	5,58
<b>32</b>	193,600	1548,51	5,56	3,86	6,74
<b>33</b>	193,650	1548,11	3,37	4,87	8,18
<b>34</b>	193,700	1547,72	2,35	5,68	9,38

<b>35</b>	193,750	1547,32	5,59	6,55	10,71
<b>36</b>	193,800	1546,92	3,35	8,01	12,9
<b>37</b>	193,850	1546,52	2,31	9,3	14,61
<b>38</b>	193,900	1546,12	5,24	11,06	16,73
<b>39</b>	193,950	1545,72	2,91	13,29	19,6
<b>40</b>	194,000	1545,32	1,59	15,79	22,58
<b>41</b>	194,050	1544,92	3,62	17,59	24,39
<b>42</b>	194,100	1544,53	1,2	19,21	26,6
<b>43</b>	194,150	1544,13	2,95	20,61	5,07
<b>44</b>	194,200	1543,73	5,31	2,29	5,61
<b>45</b>	194,250	1543,33	3,37	2,86	6,74
<b>46</b>	194,300	1542,94	1,89	3,65	8,01
<b>47</b>	194,350	1542,54	5	4,27	9,31
<b>48</b>	194,400	1542,14	3,23	5,2	11,05
<b>49</b>	194,450	1541,75	2,3	5,86	12,35
<b>50</b>	194,500	1541,35	4,89	7,25	14,47
<b>51</b>	194,550	1540,95	3,48	8,54	16,35
<b>52</b>	194,600	1540,56	2,1	9,83	18,65
<b>53</b>	194,650	1540,16	4,11	11,89	21,28
<b>54</b>	194,700	1539,77	2,37	13,76	23,67
<b>55</b>	194,750	1539,37	3,85	16,18	25,76
<b>56</b>	194,800	1538,98	2,31	16,3	4,86
<b>57</b>	194,850	1538,58	4,54	19,29	5,99
<b>58</b>	194,900	1538,19	5,63	1,94	6,81
<b>59</b>	194,950	1537,79	3,9	2,42	8,01
<b>60</b>	195,000	1537,40	2,67	2,8	8,97
<b>61</b>	195,050	1537,00	1,04	3,88	11,6
<b>62</b>	195,100	1536,61	3,37	4,57	12,73
<b>63</b>	195,150	1536,22	2,19	5,4	14,75
<b>64</b>	195,200	1535,82	1,34	6,3	16,9
<b>65</b>	195,250	1535,43	0,62	7,78	19,4
<b>66</b>	195,300	1535,04	2,12	8,75	21,35
<b>67</b>	195,350	1534,64	4,43	10,38	23,47
<b>68</b>	195,400	1534,25	5,44	11,93	4,01
<b>69</b>	195,450	1533,86	3,85	13,27	4,73
<b>70</b>	195,500	1533,47	2,19	15,83	5,75

71	195,550	1533,07	4,36	18,51	7,02
72	195,600	1532,68	1,31	1,69	8,01
73	195,650	1532,29	1,08	2,2	0,69
74	195,700	1531,90	1,86	2,75	11,15
75	195,750	1531,51	2,07	3,17	1,13
76	195,800	1531,12	3,85	4,22	1,54
77	195,850	1530,72	3,37	4,59	1,78
78	195,900	1530,33	2,26	5,29	2,16
79	195,950	1529,94	4,4	6,86	2,81
80	196,000	1529,55	2,53	8,24	3,43

Προκειμένου να βελτιωθεί περαιτέρω το επιλεγμένο μήκος κύματος ακριβώς στην κεντρική συχνότητα σύμφωνα με το ITU για καθένα από τα κανάλια DWDM, ρυθμίζεται (*fine-tuning*) ανάλογα το ρεύμα που αντιστοιχεί στο τμήμα φάσης του λέιζερ. Ωστόσο, όπως παρατηρήθηκε πειραματικά, οι τιμές ρεύματος που εγχύθηκαν στη φάση του λέιζερ δεν παρουσίαζαν την ίδια επαναληψιμότητα μεταξύ διαφορετικών μετρήσεων για το ίδιο κανάλι DWDM με αυτές που παρατηρήθηκαν για τα τμήματα ανακλαστήρα. Αντ' αυτού, κάθε φορά που έπρεπε να ρυθμιστεί κατάλληλα το ρεύμα φάσης μεταξύ 0-7 mA για να επιτευχθεί η κεντρική συχνότητα σύμφωνα με το ITU grid. Ωστόσο, δεδομένου ότι στην προτεινόμενη υβριδική αρχιτεκτονική μεταγωγής χρησιμοποιούνται συστοιχίες οπτικών φίλτρων (AWG) όπου η απόσταση μεταξύ των καναλιών είναι στα 50 GHz και το ωφέλιμο εύρος ζώνης για καθένα από τα κανάλια DWDM είναι περίπου 37 GHz, δεν είναι αυστηρή απαίτηση να επιτευχθεί η κεντρική συχνότητα.

Η μεταβλητότητα του μήκους κύματος εκπομπής λέιζερ μέσω της βελτίωσης του τμήματος φάσης του απεικονίζεται στο διάγραμμα στην Εικόνα 19 [52], που δείχνει τη σχέση του μήκους κύματος λέιζερ ως συνάρτηση του ρεύματος φάσης, υποθέτοντας ένα σταθερό ρεύμα οδήγησης για τους δύο ανακλαστήρες. Μπορούμε να συμπεράνουμε ότι η μεταβολή του ρεύματος φάσης εντός της περιοχής λειτουργίας του επιτρέπει τον συντονισμό του μήκους κύματος του λέιζερ κατά λιγότερο από 1 nm (125 GHz), το οποίο είναι σημαντικά ευρύτερο από το διάστημα καναλιών.

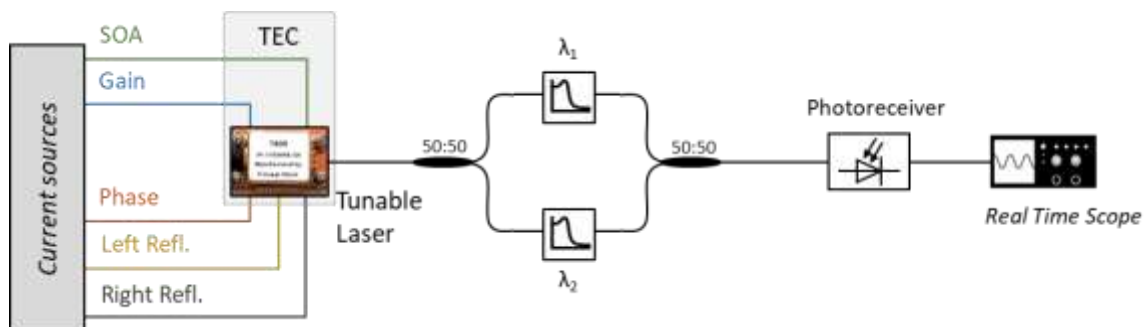


**Εικόνα 19:** Διάγραμμα που απεικονίζει την μεταβολή του μήκους κύματος σε σχέση με την μεταβολή του ρεύματος στο τμήμα της φάσης του λέιζερ. Οι τιμές στα τμήματα των ανακλαστήρων του λέιζερ έχουν ρυθμιστεί σταθερά στις τιμές 8.56 mA (αριστερός ανακλαστήρας) και 9.93 mA (δεξιός ανακλαστήρας) [52].

#### 2.1.4 Χαρακτηρισμός ρύθμισης μήκους κύματος του λέιζερ υπό δυναμική λειτουργία

Για τον προσδιορισμό του χρόνου που απαιτείται για τον συντονισμό του μήκους κύματος του λέιζερ (tuning time) ακολουθήθηκε μια διαδικασία δυναμικού χαρακτηρισμού του λέιζερ.

Η λογική που ακολουθείται σε αυτή τη μεθοδολογία είναι η εξής και απεικονίζεται στην Εικόνα 20 το λέιζερ ρυθμίζεται έτσι ώστε να μεταβαίνει από ένα μήκος κύματος  $\lambda_1$  σε ένα άλλο  $\lambda_2$  εναλλάξ και επαναλαμβανόμενα. Με τη χρήση ενός οπτικού διαχωριστή ισχύος 1×2 στην έξοδο του λέιζερ, δημιουργούνται δύο οπτικές γραμμές. Προκειμένου κάθε οπτική γραμμή να αντιστοιχεί σε ένα μήκος κύματος, χρησιμοποιούνται οπτικά φίλτρα με κεντρικές συχνότητες τις συχνότητες  $\lambda_1$  και  $\lambda_2$  αντίστοιχα, και έτσι γίνεται δυνατή η παρατήρηση ξεχωριστά των δύο κυμάτων. Αφού τα δύο οπτικά κύματα περάσουν μέσα από τα ζωνοπερατά φίλτρα εύρους 37.5GHz συνδυάζονται εκ νέου μέσω ενός οπτικού συζεύκτη 2×1 και διοχετεύονται σε έναν φωτοανιχνευτή, αναμένοντας να εμφανιστεί ένα χρονικό διάστημα κατά την μετάβαση από το ένα μήκος κύματος στο άλλο όπου δεν θα φτάνει φως στον φωτοανιχνευτή. Αυτό το χρονικό διάστημα θα αποτελεί τον χρόνο συντονισμού του λέιζερ από ένα μήκος κύματος σε ένα άλλο.

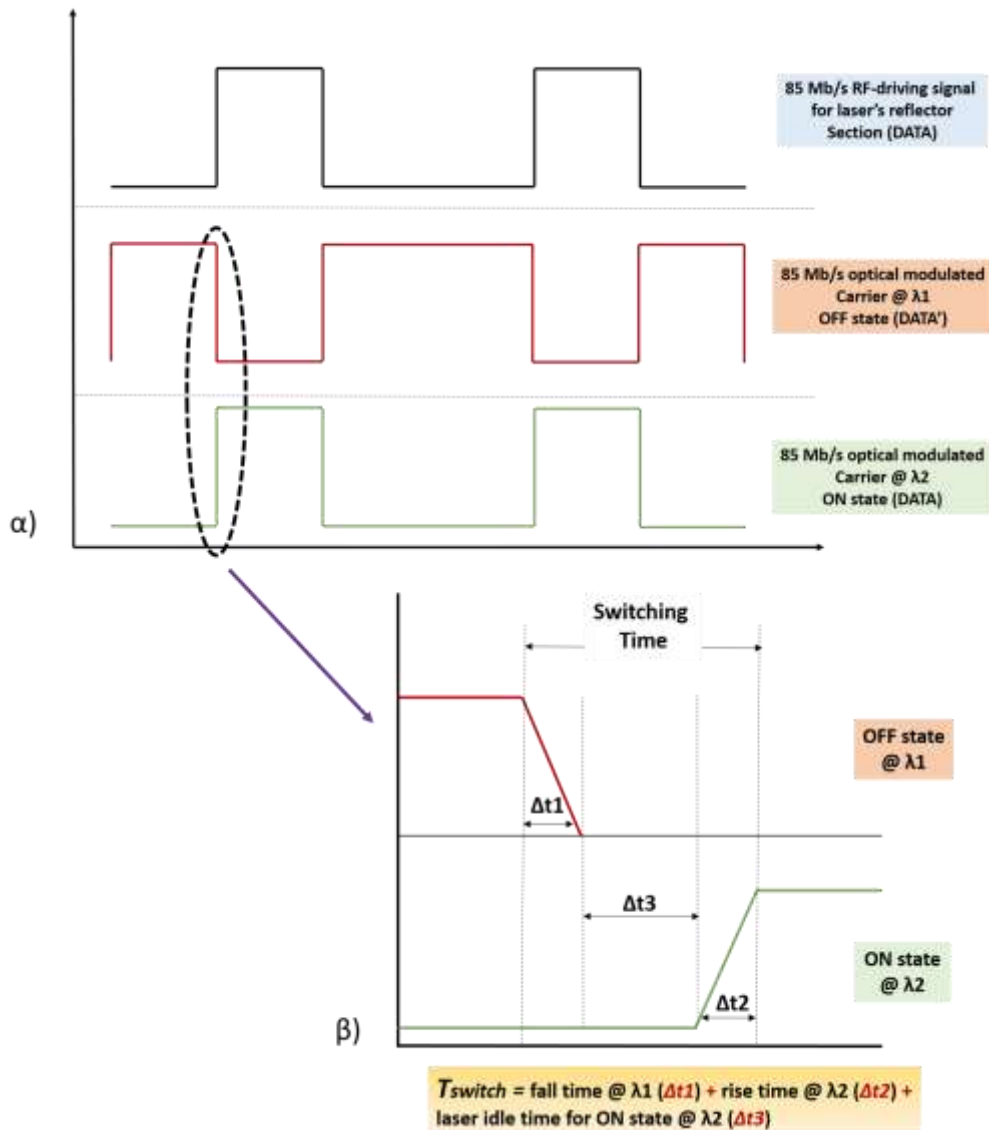


**Εικόνα 20:** Μεθοδολογία για τον προσδιορισμό του χρόνου που απαιτείται για τον συντονισμό του μήκους κύματος του λέιζερ (tuning time)



Δεδομένου ότι κατά τη στιγμή των πρώτων πειραμάτων χαρακτηρισμού, δεν ήταν διαθέσιμος ένας γρήγορος ψηφιακός μετατροπέας ρεύματος πολλαπλών καναλιών (DAC) για τον ταυτόχρονο έλεγχο των τριών ρευμάτων συντονισμού του λέιζερ (δηλαδή τα ρεύματα για τους δύο ανακλαστήρες και το τμήμα της φάσης), ο δυναμικός συντονισμός του λέιζερ επιχειρήθηκε ρυθμίζοντας μόνο μία μόνο παράμετρο, χρησιμοποιώντας μια γεννήτρια παλμών (Pulse Pattern Generator - PPG). Όπως περιγράφεται και παρακάτω, έγινε διαμόρφωση ενός από τους δύο ανακλαστήρες με NRZ ηλεκτρικό σήμα σε ρυθμό 85 Mb/s, που αντιστοιχεί σε περίοδο συντονισμού (tuning period)  $<12$  ns.

Οι πρώτοι παλμοί που φαίνονται με μαύρο χρώμα στην Εικόνα 21 (α) αντιστοιχούν σε αυτά που παράγονται από την γεννήτρια παλμών και που διαμορφώνουν τον δεξιό ανακλαστήρα του λέιζερ, σύμφωνα με την μεθοδολογία που περιγράφηκε προηγουμένως. Όταν αυτό το ψηφιακό ηλεκτρικό σήμα εισάγεται στον ανακλαστήρα του λέιζερ, στην πραγματικότητα τροποποιεί τις ιδιότητες συντονισμού του λέιζερ σύμφωνα με τα μηδενικά «0» και τους άσσους «1» του μοτίβου (μαύρο ίχνος στο σχήμα στην Εικόνα 21 (α)). Ρυθμίζοντας κατάλληλα τις υπόλοιπες παραμέτρους συντονισμού του λέιζερ (δεύτερος ανακλαστήρας, φάση, κέρδος και SOA) σε μια σταθερή τιμή, πετυχαίνουμε εναλλαγή του συντονισμού του λέιζερ μεταξύ δύο σταθερών καταστάσεων μήκους κύματος, δηλαδή στο μήκος κύματος  $\lambda_1$  και στο μήκος κύματος  $\lambda_2$ .



**Εικόνα 21:** α) Το σήμα για τον δυναμικό χαρακτηρισμό του προσαρμοζόμενου λέιζερ: Κορυφή: ηλεκτρικό σήμα οδήγησης στον ανακλαστήρα. μεσαίο: οπτικό σήμα μετά το φιλτράρισμα στο  $\lambda_1$ ; κάτω: οπτικό σήμα μετά το φιλτράρισμα στο  $\lambda_2$ . β) Σχηματική αναπαράσταση για τον προσδιορισμό του χρόνου συντονισμού του λέιζερ

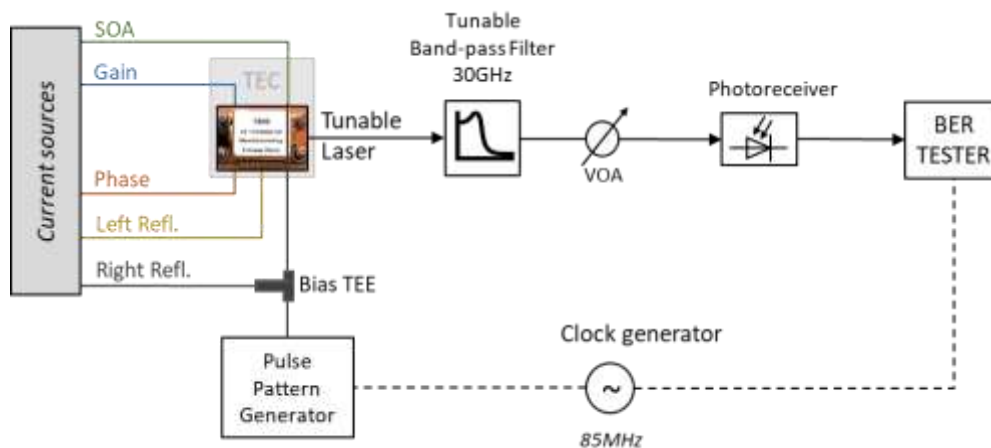
Δεδομένου ότι οι φωτοδιόδοι είναι αγνωστικές ως προς το μήκος κύματος που ανιχνεύουν, για να μετρηθεί η ταχύτητα μεταγωγής του λέιζερ, είναι απαραίτητο να χρησιμοποιηθούν οπτικά φίλτρα που να διαχωρίζουν τα μήκη κύματος σε όμοιες αλλά ξεχωριστές οπτικές διαδρομές, έτσι ώστε η εναλλαγή του μήκους κύματος που λαμβάνεται από τη φωτοδίοδο να μεταφράζεται σε εναλλαγή της στάθμης φωτορεύματος στην έξοδο της φωτοδιόδου.

Ως εκ τούτου, κατά την εναλλαγή μεταξύ αυτών των δύο μηκών κύματος, οι ενδιάμεσες καταστάσεις του λέιζερ «φιλτράρονται» από το οπτικό φίλτρο και έτσι δημιουργούνται σήματα On-Off-Keying (OOK) σε μήκη κύματος  $\lambda_1$  και  $\lambda_2$  αντίστοιχα, όπως φαίνεται στο κόκκινο και πράσινο ίχνος του στην Εικόνα 21 (α). Το ψηφιακό μοτίβο που χρησιμοποιείται για την εναλλαγή των ρευμάτων του ανακλαστήρα του λέιζερ είναι αποτυπωμένο στο  $\lambda_2$ , ενώ το  $\lambda_1$  διαμορφώνεται στο συμπληρωματικό μοτίβο.

Όπως φαίνεται στο σχήμα στην Εικόνα 21 (β), κατά την εναλλαγή του μήκους κύματος εκπομπής του λέιζερ από την  $\lambda_1$  σε  $\lambda_2$  διακρίνονται τρία χρονικά διαστήματα, δηλαδή: i) ο «χρόνος πτώσης» ( $\Delta t1$ ) κατά το οποίο η οπτική ισχύς που λαμβάνεται μετά από το φίλτρο  $\lambda_1$  μειώνεται κι αυτό οφείλεται στο γεγονός ότι το μήκος κύματος του λέιζερ απομακρύνεται από το  $\lambda_1$ , ii) ο «χρόνος ανόδου» ( $\Delta t2$ ) που αφορά στο διάστημα κατά το οποίο το μήκος κύματος του λέιζερ πλησιάζει το  $\lambda_2$ , κι έτσι παρατηρείται αύξηση της οπτικής ισχύος μετά από το φίλτρο  $\lambda_2$ , iii) τον «χρόνο αδράνειας», όταν η έξοδος του λέιζερ εκπέμπει έξω από το εύρος ζώνης των δύο φίλτρων που τοποθετούνται στα  $\lambda_1$  και  $\lambda_2$ . Έτσι, ο συνολικός χρόνος μεταγωγής του λέιζερ δίνεται από το παρακάτω άθροισμα:

$$T_{switch} = \Delta t1 + \Delta t2 + \Delta t3, \quad (1)$$

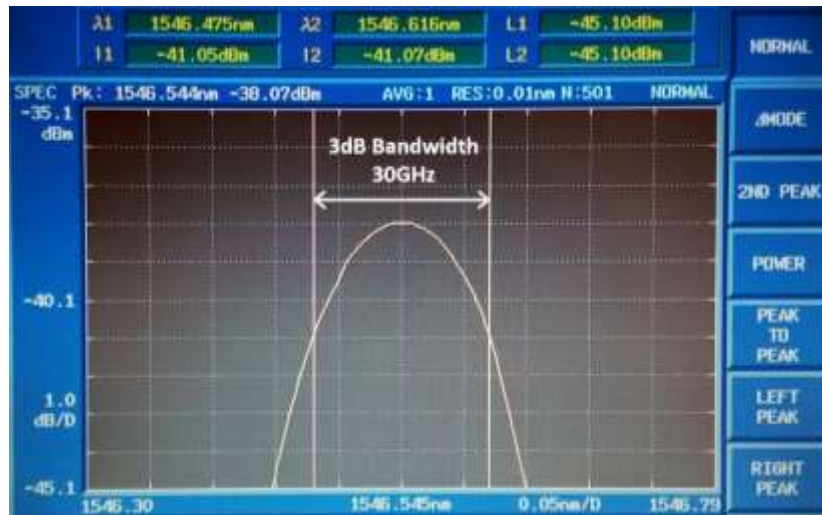
Η Εικόνα 22 απεικονίζει την πειραματική διάταξη που χρησιμοποιήθηκε για τον υπολογισμό του χρόνου εναλλαγής του προσαρμοζόμενου λέιζερ, σύμφωνα με την μεθοδολογία που περιεγράφηκε παραπάνω. Και σε αυτήν την περίπτωση, τα ρεύματα που απαιτούνται για την οδήγηση του λέιζερ παρέχονται από μια μονάδα 16 ανεξάρτητων καναλιών ρεύματος και πιο συγκεκριμένα τα ρεύματα στα τμήματα ενίσχυσης SOA και GAIN ρυθμίστηκαν στα 40 mA, δίνοντας σταθερή οπτική ισχύ εξόδου 8 dBm για κάθε ένα από τα μεταβαλλόμενα μήκη κύματος. Καθ' όλη τη διάρκεια του πειράματος η θερμοκρασία του λέιζερ ρυθμίστηκε σταθερά στους 25°C μέσω του ελεγκτή θερμοκρασίας που είναι ενσωματωμένος στο λέιζερ (TEC).



**Εικόνα 22:** Η πειραματική διάταξη που χρησιμοποιήθηκε για το δυναμικό χαρακτηρισμό του λέιζερ.

Στο συγκεκριμένο πείραμα χρησιμοποιήθηκε μόνο ένας βαθμός ελευθερίας για το συντονισμό του λέιζερ και αυτός ήταν η μεταβολή στο ρεύμα του αριστερού ανακλαστήρα. Πιο συγκεκριμένα, το ρεύμα του αριστερού ανακλαστήρα μεταβάλλεται ύστερα από υπέρθεση ενός συνεχούς ρεύματος (DC) στα 9,37 mA (που δίνεται και πάλι μέσω του ILX) και ενός μεταβαλλόμενου ρεύματος με διακύμανση από 0 έως 365 mV<sub>pp</sub> που παρέχεται από τη γεννήτρια παλμών (PPG) με ρυθμό 85 Mb/s. Αυτές οι δύο πηγές ρεύματος συνδυάστηκαν μέσω ενός bias-TEE και κατόπιν εισήχθησαν στον αριστερό ανακλαστήρα του λέιζερ μέσω της ειδικής SMA υποδοχής της πλακέτας του λέιζερ. Η εναλλαγή τάσης του σήματος που παράγει η γεννήτρια παλμών επιλέχθηκε με σκοπό η εναλλαγή του λέιζερ να γίνεται μεταξύ δύο σταθερών καταστάσεων μήκους κύματος, δηλαδή λ<sub>1</sub> στα 1527,99 nm και λ<sub>2</sub> στα 1546,52 nm, αμφότερα με SMSR > 40 dB. Παράλληλα το ρεύμα στον δεξιό ανακλαστήρα ρυθμίστηκε να είναι σταθερό στα 9,51 mA.

Δεδομένου ότι κατά την εκτέλεση του πειράματος δεν υπήρχε διαθέσιμο πλέγμα κυματοδηγών (Arrayed waveguide gratings AWG) για την αποπολυπλεξία των δύο μηκών κύματος κατά τα πρότυπα του ITU, χρησιμοποιήθηκαν οπτικά ζωνοπερατά φίλτρα (Optical bandpass filter - OBF) με μεταβλητό μήκος κύματος και εύρος ζώνης [53]. Προκειμένου να εξομοιωθεί με τα χαρακτηριστικά του AWG που θα χρησιμοποιηθούν κατά την υλοποίηση της προτεινόμενης αρχιτεκτονικής, το εύρος ζώνης 3 dB του OBF ρυθμίστηκε στα 30 GHz, και μετρήθηκε μέσω του αναλυτή οπτικού φάσματος (Optical Spectrum Analyzer – OSA). Το φάσμα του οπτικού φίλτρου που καταγράφηκε από τον αναλυτή φάσματος φαίνεται στην Εικόνα 23, αφού εισάχθηκε οπτικός θόρυβος μέσω ενός ενισχυτή ινας ερβίου (EDFA).



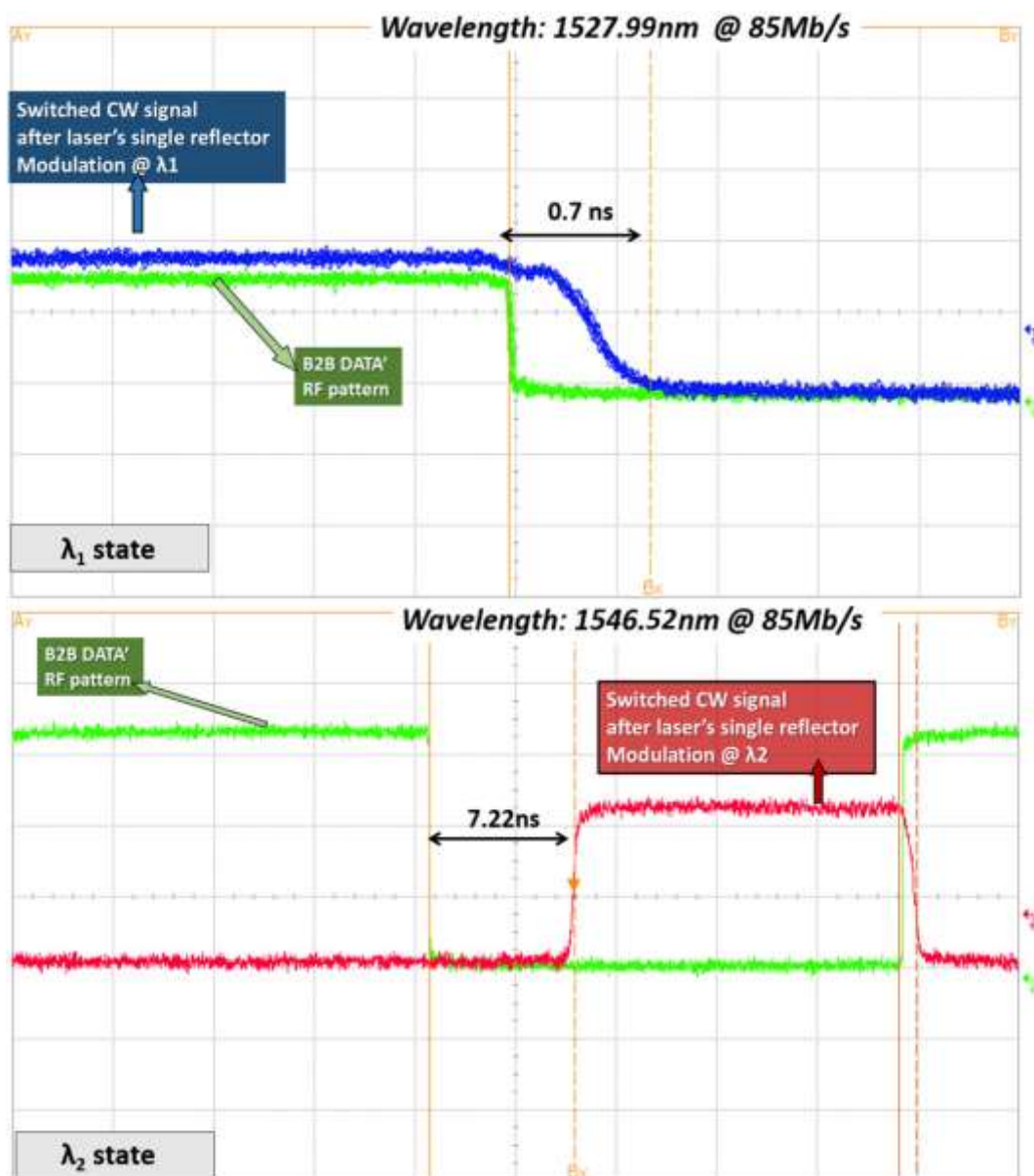
**Εικόνα 23:** Το εύρος ζώνης 3 dB του οπτικού ζωνοπερατού φίλτρου (Optical bandpass filter - OBF) με μεταβλητό μήκος κύματος και εύρος ζώνης, όπως απεικονίζεται σε ένα οπτικό φασματογράφο.

Η Εικόνα 24 απεικονίζει κυματομορφές που καταγράφηκαν από παλμογράφο σε πραγματικό χρόνο και οι οποίες δείχνουν την απόκριση χρόνου συντονισμού του λέιζερ, όταν ο αριστερός ανακλαστήρας του διαμορφώνεται με σήμα 85Mb/s. Η μετάβαση από το μήκος κύματος  $\lambda_1$  εμφανίζεται στο επάνω ίχνος μετά από οπτικό φιλτράρισμα γύρω από το  $\lambda_1$ , ενώ η μετάβαση στο  $\lambda_2$  εμφανίζεται στο κάτω ίχνος, μετά από οπτικό φιλτράρισμα γύρω από το  $\lambda_2$ . Το ηλεκτρικό σήμα οδήγησης εμφανίζεται για σύγκριση των μεταβάσεων (πράσινο ίχνος). Όπως ήταν αναμενόμενο, τα μοτίβα δεδομένων της κατάστασης  $\lambda_1$  (μπλε ίχνος) και της  $\lambda_2$  κατάστασης (κόκκινη κατάσταση) ήταν συμπληρωματικά μεταξύ τους και ένα από αυτά συνέπεσε με το μοτίβο οδήγησης που εισήχθη στον ανακλαστήρα του λέιζερ. Σύμφωνα με την εξίσωση (1) ο χρόνος μεταγωγής του λέιζερ μετρήθηκε σε  $\sim 8$  ns. Αυτό το αποτέλεσμα συμφωνεί και με τη σχετική βιβλιογραφία. Πιο συγκεκριμένα, η αναφορά [46] αναφέρει τα δυναμικά χαρακτηριστικά συντονισμού του λέιζερ MG-Y, για διαφορετικούς ρυθμούς μεταβάσεων μήκους κύματος:

1. εντός του ίδιου ρυθμού,
2. εντός του ίδιου υπερ-ρυθμού σε διαδοχικούς ρυθμούς,
3. εντός του ίδιου υπερ-ρυθμού σε μη διαδοχικούς ρυθμούς,
4. σε διαδοχικούς υπερ-ρυθμούς,
5. σε μη διαδοχικούς υπερ-ρυθμούς

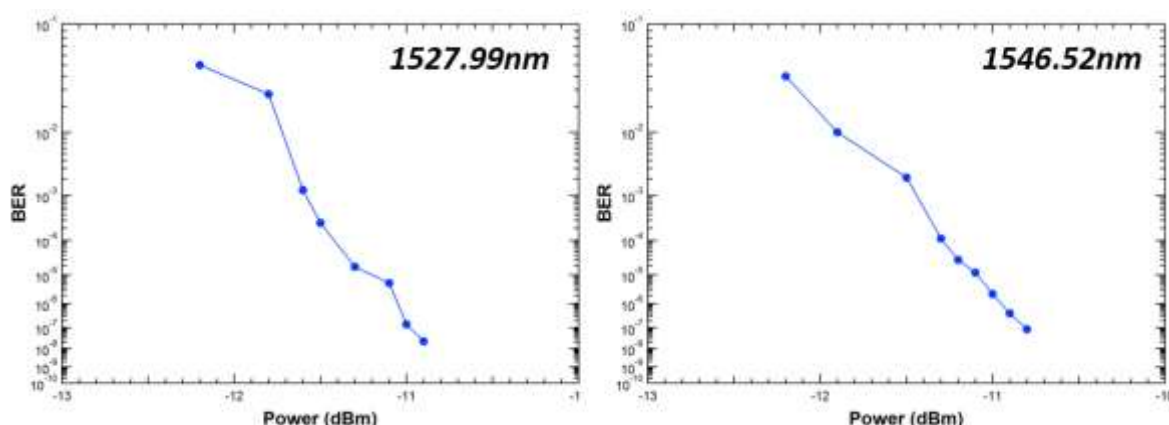
Εκτός από την περίπτωση 3 όπου το λέιζερ επιτυγχάνει χρόνο εναλλαγής κάτω από 1 ns, οι χρόνοι εναλλαγής κυμαίνονταν από 4 ns έως 13 ns. Κατά την εναλλαγή μεταξύ μη

διαδοχικών supermodes (όπως στο σενάριο που επαληθεύτηκε πειραματικά και φαίνεται στην Εικόνα 24), η αναφορά [54] αναφέρει χρόνους εναλλαγής μεταξύ 5 ns σε 13 ns, σε συμφωνία δηλαδή με την πειραματική μέτρηση των 8 ns, επιβεβαιώνοντας την επιτυχημένη εφαρμογή του PCB.



**Εικόνα 24:** Κοντινή απεικόνιση της μετάβασης του λέιζερ διαμορφωμένο στα 85Mb/s. Το λέιζερ εναλλάσσει την έξοδο από την κατάσταση λ<sub>1</sub> στα 1527,99nm (μπλε ίχνος, πάνω) στην κατάσταση λ<sub>2</sub> στα 1546,52nm (κόκκινο ίχνος, κάτω). Το πράσινο ίχνος δείχνει το ηλεκτρικό μοτίβο που οδηγεί τον αριστερό ανακλαστήρα.

Προκειμένου να γίνει ορθή αξιολόγηση της ποιότητας του οπτικού σήματος εξόδου του λέιζερ υπό δυναμική λειτουργία, καταγράφηκαν μετρήσεις διαμορφωμένου σήματος για κάθε ένα από τα δύο εναλλασσόμενα μήκη κύματος ξεχωριστά. Με αυτές τις μετρήσεις έγινε απεικόνιση των καμπύλων του ρυθμού σφάλματος bit (Bit Error Rate – BER) έναντι της λαμβανόμενης οπτικής ισχύος και για τα δύο μοτίβα που παράγονται σε  $\lambda_1$  και  $\lambda_2$  μετά από το οπτικό φιλτράρισμα. Όπως φαίνεται στα διαγράμματα στην Εικόνα 25, και τα δύο μήκη κύματος, που έχουν φασματική απόσταση  $> 18,5$  nm, εμφανίζουν απόδοση χωρίς σφάλματα και ίδια συμπεριφορά στα ίδια λαμβανόμενα επίπεδα ισχύος, επιβεβαιώνοντας ότι το λέιζερ είναι ικανό να εναλλάσσει το μήκος κύματος του σε ταχύτητα σημαντικά υψηλή, διατηρώντας την ακεραιότητα του σήματος.



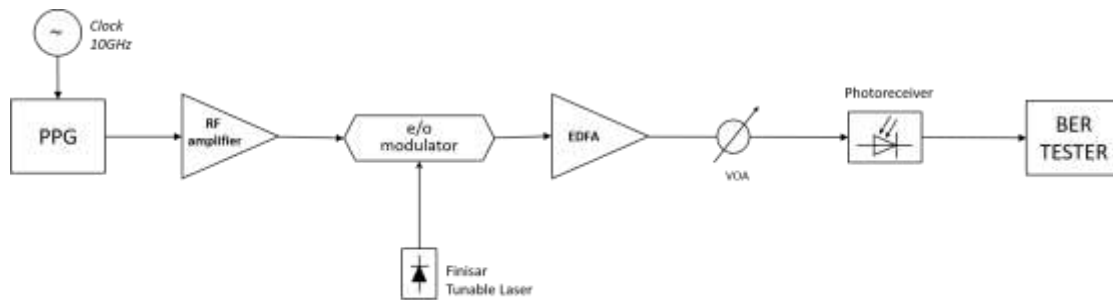
**Εικόνα 25:** Μετρήσεις bit-error-rate (BER) σε σχέση με την λαμβανόμενη οπτική ισχύ στον δέκτη για δύο εναλλασσόμενα μήκη κύματος υπό δυναμικό έλεγχο του λέιζερ στα 85Mb/s.

Λεπτομερής χαρακτηρισμός του προσαρμοζόμενου λέιζερ, όταν αυτό ενσωματώνεται στον πομπό, περιγράφεται στην ενότητα 3.2.2. Σε αυτήν την ενότητα προκειμένου να γίνει η δυναμική ρύθμιση των ρευμάτων με υψηλή ακρίβεια, χρησιμοποιείται ένας μετατροπέας ψηφιακού σε αναλογικό (digital-to-analog converter DAC) 4 καναλιών. Ο DAC (Texas Instruments DAC34H84EVM [55]) προγραμματίζεται μέσω μιας συστοιχίας επιτόπια προγραμματιζόμενων πυλών (Field Programmable Gate Array – FPGA).

### 2.1.5 Εξωτερική διαμόρφωση του προσαρμοζόμενου λέιζερ στα 10Gb/s με παράλληλη αξιολόγηση του πομπού.

Στα πλαίσια της ανάπτυξης του προσαρμοζόμενου πομπού, η διάταξη του προσαρμοζόμενου λέιζερ που περιγράφεται στις προηγούμενες ενότητες συνδυάστηκε με έναν εξωτερικό οπτικό διαμορφωτή, ενώ έγινε και αξιολόγηση της απόδοσής του στα 10 Gb/s σε παλμούς Non-Return-Zero (NRZ).

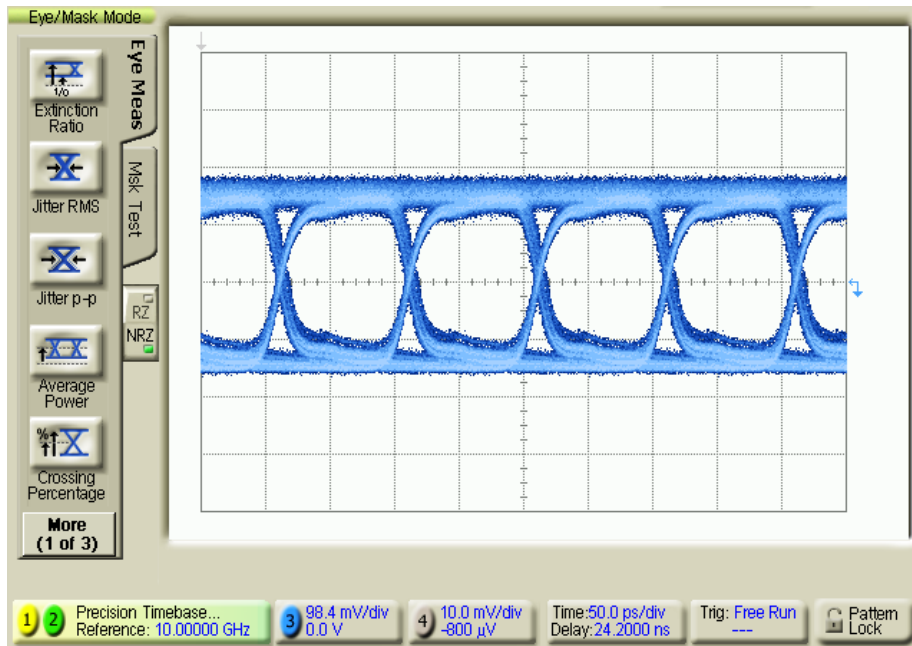
Η Εικόνα 26 απεικονίζει την πειραματική διάταξη που χρησιμοποιήθηκε για την αξιολόγηση του συντονισμένου λέιζερ υπό εξωτερική διαμόρφωση υψηλής ταχύτητας στα 10 Gb/s. Ο εξωτερικός ηλεκτρο-οπτικός διαμορφωτής ήταν ένας εμπορικός MZM με εύρος ζώνης ~25 GHz στα 3dB (3dB bandwidth) σε single-drive λειτουργία. Το υπό διαμόρφωση ηλεκτρικό σήμα ήταν μια ψευδοτυχαία ακολουθία από  $10^{27-1}$  bit στα 10GHz (PRBS) που παρέχεται από μια γεννήτρια παλμών (Pulse Pattern Generator), μετά από ενίσχυση σε έναν ευρυζωνικό ενισχυτή ραδιοσυχνοτήτων (RF amplifier), για να ταιριάζει με την τάση  $V_{\pi}$  του διαμορφωτή στα 3Vpp. Η διαμορφωμένη έξοδος του MZM εισήχθη εν συνεχεία σε έναν οπτικό ενισχυτή ίνας ερβίου (EDFA) και έναν μεταβλητό οπτικό εξασθενητή (VOA) ακολουθούμενο από έναν εμπορικό φωτοδέκτη 40 Gb/s, προκειμένου να αξιολογηθεί η απόδοση BER του πομπού ως συνάρτηση της λαμβανόμενης οπτικής ισχύος. Προκειμένου να εκτιμηθεί οποιαδήποτε πιθανή εξάρτηση από το μήκος κύματος της απόδοσης του πομπού (π.χ. λόγω παραλλαγών ισχύος ή OSNR του λέιζερ σε όλη τη ζώνη C), μετρήθηκαν τρία κανάλια, δηλαδή: 1561,01 nm (channel 1), 1529,55 nm (channel 80) και 1554,94 nm (channel 16).



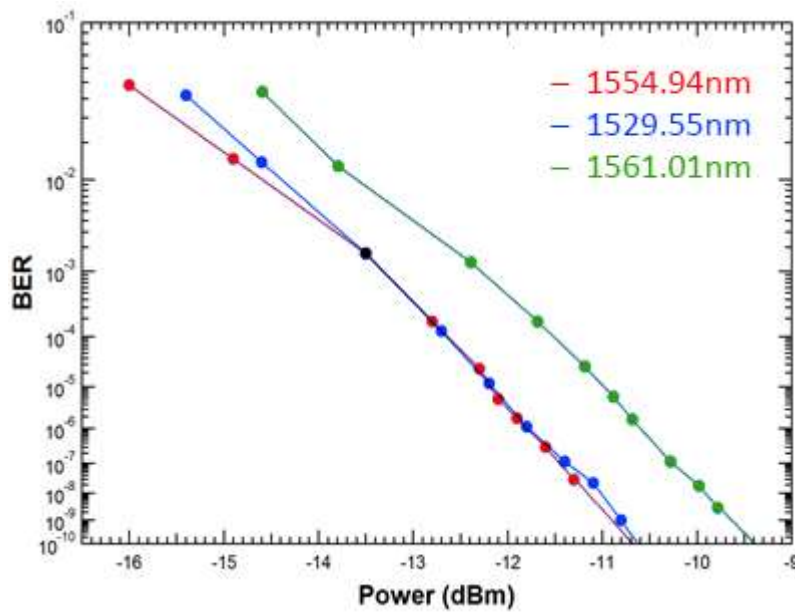
**Εικόνα 26:** Η πειραματική διάταξη που χρησιμοποιήθηκε για την αξιολόγηση του προσαρμοζόμενου πομπού στα 10Gb/s.

Η Εικόνα 27 δείχνει ένα τυπικό οπτικό διάγραμμα ματιού (eye-diagram) που προήλθε από τον πομπό με ρυθμό μετάδοσης 10 Gb/s εγγεγραμμένο σε μήκος κύματος 1529,55 nm. Δεν παρατηρήθηκε διακριτή παραλλαγή στα διαγράμματα ματιών των άλλων δύο μηκών κύματος. Το διάγραμμα στην Εικόνα 28 δείχνει τις καμπύλες BER και για τα τρία διαμορφωμένα μήκη κύματος. Λειτουργία χωρίς σφάλματα (επίπεδο  $<10^{-10}$  BER) επιτεύχθηκε για όλα τα μήκη κύματος με το μήκος κύματος 1561.01nm να εμφανίζει ελαφριά ποινη ισχύος ~ 1dB σε σχέση με το 1529,55nm. Το τελευταίο μπορεί να εξηγηθεί από το ανομοιογενές φάσμα κέρδους του οπτικού ενισχυτή και τα τμήματα κέρδους εντός του προσαρμοζόμενου λέιζερ.





*Εικόνα 27: Στιγμιότυπο από το διάγραμμα ματιού (eye-diagram) του προσαρμοζόμενου πομπού στα 10Gb/s*



*Εικόνα 28: Καμπύλες BER σε σχέση με την λαμβανόμενη ισχύ για διαμόρφωση 10Gb/s για μήκη κύματος 1529.55 nm (ch.80), 1561.01 nm (ch.1) και 1554.94 nm (ch.16) που παράγονται από τον συντονιζόμενο πομπό.*

## 2.2 Ανάπτυξη και διαμόρφωση του δέκτη του μεταγωγέα ToR switch της προτεινόμενης υβριδικής αρχιτεκτονικής

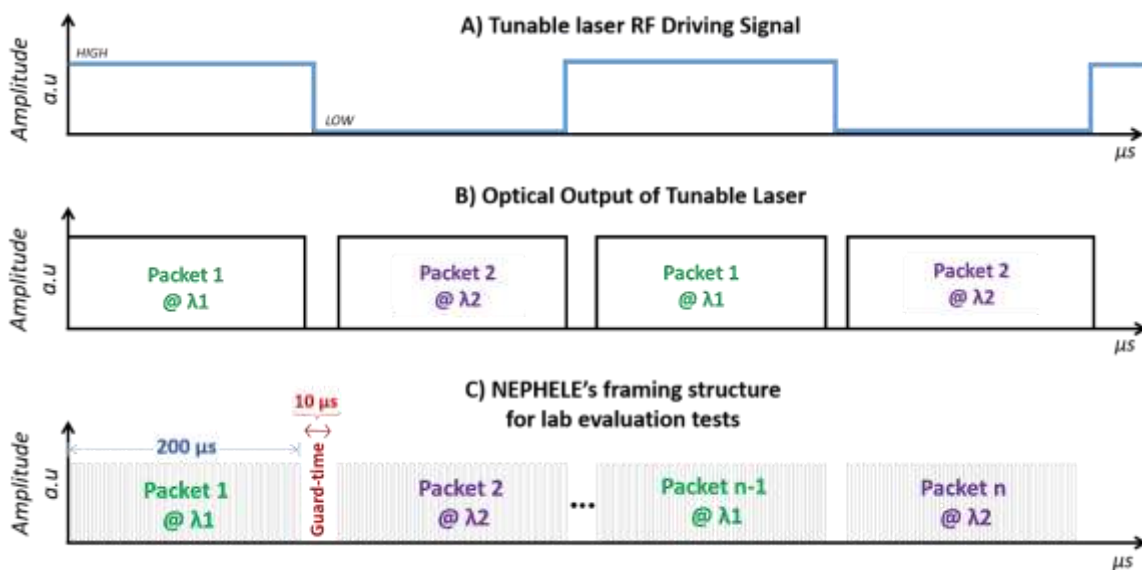
Ο μεταγωγέας του ToR της προτεινόμενης υβριδικής αρχιτεκτονικής, από εδώ και πέρα για ευκολία θα αναφέρεται ως *ToR switch*, βασίζεται σε έναν πομπό με δυνατότητα ρύθμισης μήκους κύματος, ο οποίος είναι υπεύθυνος για την καταγραφή των ηλεκτρικών δεδομένων (που προέρχονται από τους κεντρικούς υπολογιστές Ethernet) σε οπτικό φορέα σε ρυθμό 10Gb/s. Σύμφωνα με την Αρχή λειτουργίας της μελετώμενης αρχιτεκτονικής, η ίδια η οπτική συχνότητα του πομπού καθορίζει και τη δρομολόγηση των δεδομένων προς τον κατάλληλο προορισμό. Η ανάπτυξη του προσαρμοζόμενου πομπού 10 Gb/s αναφέρθηκε στην ενότητα 2.1, ενώ η ταχύτητα συντονισμού του μήκους κύματος του προσαρμοζόμενου λέιζερ μετρήθηκε να είναι πολύ πιο γρήγορη από την απαιτούμενη για την υπό διερεύνηση υβριδική αρχιτεκτονική. Πιο συγκεκριμένα, η μέγιστη ταχύτητα συντονισμού του προσαρμοζόμενου λέιζερ (δηλαδή η ικανότητά του να αλλάζει το μήκος κύματος του) έφτασε τα 80 MHz. Δεδομένου ότι το ToR switch δεν μπορεί να μεταδώσει δεδομένα κατά τη διάρκεια της ρύθμισης του λέιζερ σε διαφορετικό μήκος κύματος, ο κενός αυτός χρόνος είναι ωφέλιμος για την βελτιστοποίηση της χρήσης των πόρων του δικτύου της προτεινόμενης υβριδικής αρχιτεκτονικής. Παρόλα αυτά, η εναλλαγή μεταξύ μήκους κύματος σε τόσο υψηλό ρυθμό δεν είναι απαραίτητη στην προτεινόμενη υβριδική αρχιτεκτονική, καθώς οι προδιαγραφές καθορίζουν το μήκος των οπτικών πακέτων του και των χρονοθυρίδων («slots») σε κάποια μικροδευτερολέπτα (us). Έτσι λοιπόν, ήταν απαραίτητο να ρυθμιστεί ο πομπός ώστε να μπορεί να διαμορφώνει πακέτα με χαρακτηριστικά όμοια με εκείνα των πακέτων της προτεινόμενης υβριδικής αρχιτεκτονικής, έτσι ώστε να επαληθευθεί η επιτυχής λειτουργία στο επίπεδο των δεδομένων.



**Εικόνα 29:** Η δομή των πακέτων της αρχιτεκτονικής του προτεινόμενου υβριδικού δικτύου

Η δομή των πακέτων της προτεινόμενης υβριδικής αρχιτεκτονικής φαίνεται στην Εικόνα 29. Η διάρκεια των πακέτων είναι σταθερή στα 200  $\mu\text{s}$ , ακολουθούμενη από μια περίοδο παύσης 10  $\mu\text{s}$  (guard-time). Αυτό το μοτίβο των πακέτων που περιγράφεται στην Εικόνα 29 ακολουθείται σε κάθε χρονική στιγμή για όλες τις επικοινωνίες εντός του οπτικού δικτύου.

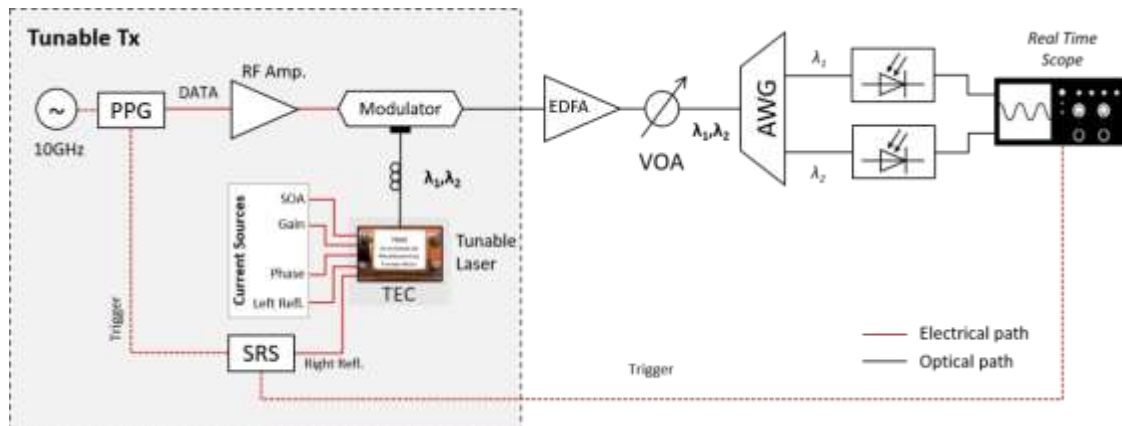
Όπως έχει αναφερθεί και σε προηγούμενες ενότητες, για τη δυναμική εναλλαγή του λέιζερ μεταξύ δύο καταστάσεων μήκους κύματος (λόγου χάρη  $\lambda_1$  και  $\lambda_2$ ), είναι απαραίτητο ένα ηλεκτρικό σήμα ελέγχου για τουλάχιστον ένα από τα δύο τμήματα ανακλαστήρων του λέιζερ. Στα επόμενα πειράματα που παρουσιάζονται, το σήμα ελέγχου εφαρμόζεται μόνο στον δεξιό ανακλαστήρα και αποτελείται από δύο στάθμες ελέγχου HIGH και LOW, όπως φαίνονται στην Εικόνα 30 (α). Μέσω αυτών των δύο σταθμών καθορίζεται αφενός η συχνότητα εκπομπής του λέιζερ και αφετέρου η διάρκεια εκπομπής σε κάθε μία κατάσταση  $\lambda_1$  και  $\lambda_2$ . Πιο συγκεκριμένα, όπως φαίνεται και στην Εικόνα 30 (β) ο εκπεμπόμενος οπτικός φορέας του λέιζερ εναλλάσσεται μεταξύ δύο μηκών κύματος  $\lambda_1$  και  $\lambda_2$  σύμφωνα με το σήμα ελέγχου κυμαινόμενο από το επίπεδο LOW που διαμορφώνει το λέιζερ ώστε να εκπέμπει στο  $\lambda_1$  και το επίπεδο HIGH που το διαμορφώνει ώστε να εκπέμπει στο  $\lambda_2$ . Πρέπει να σημειωθεί ότι για τα πειράματα που περιγράφονται στην ενότητα αυτή, χρησιμοποιήθηκαν δύο μήκη κύματος, κάνοντας προσομοίωση λειτουργίας του προτεινόμενου δικτύου με 2 ToRs ανά rod), λόγω πρακτικών περιορισμών που σχετίζονται με τον σύγχρονο συντονισμό του λέιζερ, αφού δεν ήταν διαθέσιμος ψηφιακός τρόπος παραγωγής πολλαπλών σταθμών για περαιτέρω ρύθμιση.



**Εικόνα 30:** α) Σήμα ελέγχου του δεξιού ανακλαστήρα του λέιζερ, β) Η έξοδος του προσαρμοζόμενου λέιζερ που εναλλάσσεται μεταξύ δύο μηκών κύματος ανάλογα με το σήμα ελέγχου που φαίνεται στο α), γ) διαμορφωμένα πακέτα σε ρυθμό 10Gb/s που παράγονται από τον προσαρμοζόμενο πομπό.

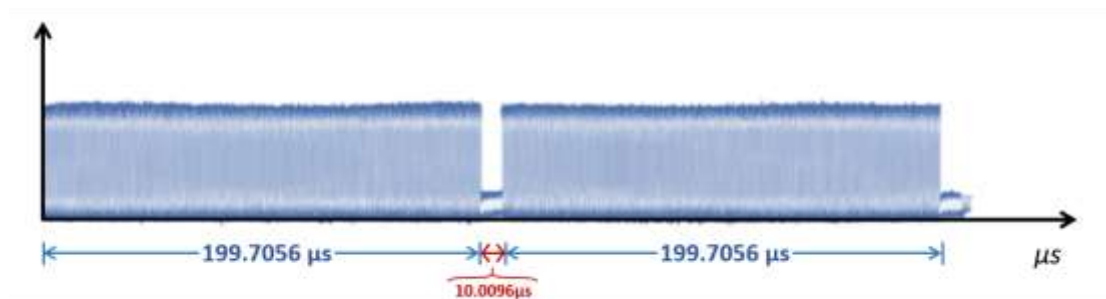
Ένα σύνολο πειραμάτων πραγματοποιήθηκε προκειμένου να διαμορφωθεί ο πομπός του ToR switch έτσι ώστε να δημιουργηθούν πακέτα δεδομένων ακολουθώντας την τυπική δομή των πακέτων της προτεινόμενης υβριδικής αρχιτεκτονικής, όπως φαίνεται στην Εικόνα 29. Η πειραματική ρύθμιση για τις δοκιμές φαίνεται στην Εικόνα 31. Ο πομπός

ακολουθεί τη διάταξη που περιγράφεται και χαρακτηρίζεται στην ενότητα 2.1, όπου το προσαρμοζόμενο λέιζερ που ακολουθείται από έναν MZM διαμορφωτή LiNbO<sub>3</sub> 10GHz και δημιουργεί πακέτα δεδομένων και τα εγγράφει τον παραγόμενο οπτικό φορέα.



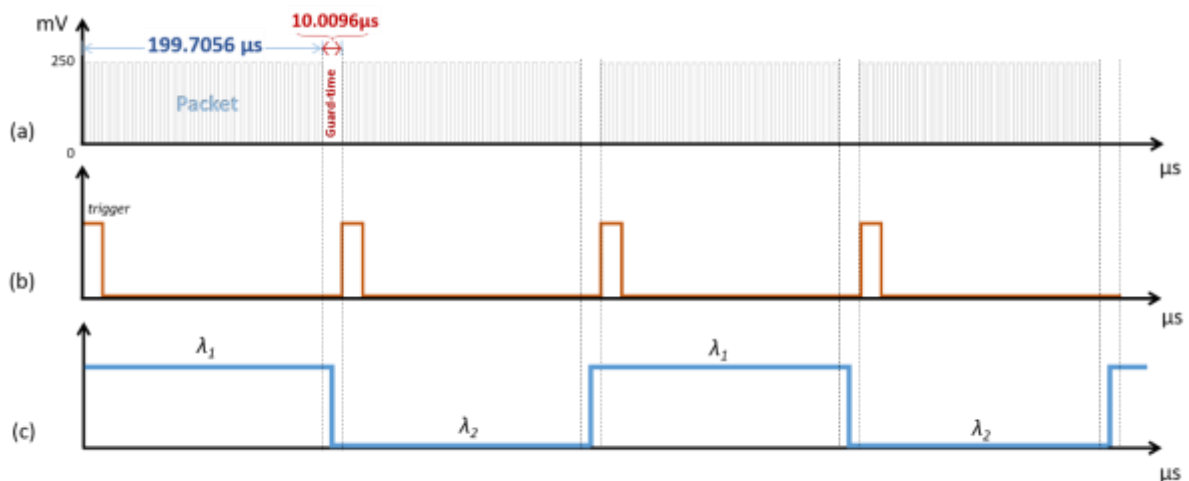
**Εικόνα 31:** Η πειραματική διάταξη για την δημιουργία και αξιολόγηση των 10Gb/s πακέτων της υβριδικής αρχιτεκτονικής.

Τα πακέτα δεδομένων δημιουργούνται από μια γεννήτρια μοτίβων παλμών (Anritsu 10 Gb/s Pulse Pattern Generator) που προγραμματίστηκε κατάλληλα ώστε να δημιουργήσει ένα επαναλαμβανόμενο μοτίβο δεδομένων βασισμένο σε μια ψευδοτυχαία ακολουθία  $2^7-1$  (PRBS) που να προσομοιάζει πακέτα. Ακολουθίες διαδοχικών μηδενικών εισήχθησαν στην PRBS για τη δημιουργία του χρόνου παύσης μεταξύ των χρονοθυρίδων-πακέτων. Το παραγόμενο σήμα των πακέτων με τα δεδομένα στην έξοδο του PPG φαίνεται στην Εικόνα 32, δείχνοντας δύο διαδοχικά πακέτα και τον σχετικό χρόνο παύσης ανάμεσά τους. Κατά την πειραματική εφαρμογή, η χρονική διάρκεια των πακέτων δεδομένων που δημιουργήθηκαν ήταν 199.7056  $\mu\text{s}$  ενώ οι περίοδοι παύσης ήταν 10.0096  $\mu\text{s}$ , έτσι ώστε να δημιουργηθεί ένα περιοδικό σήμα που μπορεί να εμφανιστεί επιτυχημένα στον παλμογράφο ισοδύναμου χρόνου και που μοιάζει πολύ με τις προδιαγραφές της υβριδικής αρχιτεκτονικής. Ακολουθώντας τα παραπάνω, τα πακέτα δεδομένων που δημιουργήθηκαν αποτελούνταν από 15602 πλήρεις περιόδους μιας  $2^7-1$  ψευδοακολουθίας PRBS.



**Εικόνα 32:** Περιοδικό μοτίβο δεδομένων στα 10Gb/s που παράγονται από τον PPG, προσομοιάζοντας τα πακέτα και το χρόνο παύσης της υβριδικής αρχιτεκτονικής.

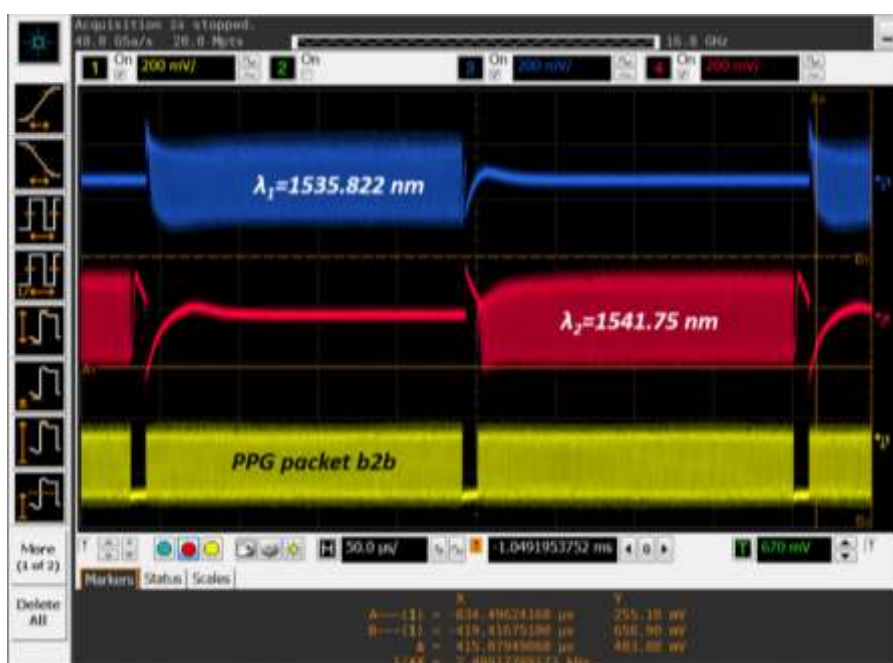
Η γεννήτρια μοτίβων δεδομένων (PPG), εκτός από τα πακέτα των 10Gb/s, παράγει κι ένα σήμα συγχρονισμού με περιοδικότητα ίση με εκείνη των πακέτων δεδομένων, δηλαδή  $1 / (199.7056 \mu\text{s} + 10.0096 \mu\text{s}) \approx 4.768 \text{ kHz}$ , όπως φαίνεται στην Εικόνα 33(β). Αυτό το σήμα συγχρονισμού «κλειδώνει» μια προγραμματιζόμενη γεννήτρια σημάτων καθυστέρησης SRS DG535 (PDG) η οποία ρυθμίστηκε έτσι ώστε να παράγει παλμούς με κύκλο λειτουργίας 50%, στο ήμισυ της περιόδου του σήματος συγχρονισμού, δηλαδή 2,384 kHz, όπως φαίνεται στην Εικόνα 33(γ). Οι παλμοί δύο σταθμών που παράγονται από την προγραμματιζόμενη γεννήτρια καθυστέρησης SRS χρησιμοποιήθηκαν για να ελέγξουν τον δεξιό ανακλαστήρα του λέιζερ και έτσι επιτεύχθηκε η ταυτόχρονη εναλλαγή του μήκους κύματος μεταξύ  $\lambda_1 = 1535.822 \text{ nm}$  και  $\lambda_2 = 1541.75 \text{ nm}$  με την παραγωγή πακέτων. Λόγω της δομής των σημάτων, τα τμήματα HIGH και LOW των παλμών που δημιουργήθηκαν ήταν ελαφρώς μεγαλύτερα από το μήκος του πακέτου, διασφαλίζοντας ότι ο χρόνος ρύθμισης του λέιζερ (χρόνοι αύξησης και πτώσης) γίνεται κατά τη διάρκεια της περιόδου παύσης και όχι κατά τη διάρκεια που μεταδίδονται τα πακέτα δεδομένων.



**Εικόνα 33:** α) Τα πακέτα δεδομένων σε ρυθμό 10 Gb/s που παράγονται από τον PPG, β) το σήμα «κλειδώματος» που προέρχεται από τον PPG και χρησιμεύει ως σήμα συγχρονισμού για την προγραμματιζόμενη γεννήτρια σημάτων καθυστέρησης (SRS), γ) Το ηλεκτρικό σήμα δύο σταθμών που παράγεται από την SRS και χρησιμεύει στον έλεγχο του δεξιού ανακλαστήρα του λέιζερ, ώστε να εκπέμπει στα μήκη κύματος  $\lambda_1=1535.822 \text{ nm}$  and  $\lambda_2=1541.75 \text{ nm}$ .

Όπως περιεγράφηκε παραπάνω, το λέιζερ εκπέμπει εναλλάξ δύο μήκη κύματος με διάρκεια εναλλαγής όσο ένα πακέτο ενώ κάθε μήκος κύματος αντιστοιχεί σε διαφορετικό πακέτο. Οι

δύο εναλλασσόμενοι οπτικοί φορείς εν συνεχεία διαμορφώνονται μέσω ενός Mach Zehnder 10GHz (MZM) διαμορφωτή με τα πακέτα δεδομένων που παράγονται από τον PPG, αφού πρώτα ενισχύονται ηλεκτρικά μέσω ενός RF ενισχυτή ευρείας ζώνης (Εικόνα 31). Τα παραγόμενα πακέτα 10 Gb/s που πλέον εγγράφονται εναλλάξ στους οπτικούς φορείς αρχικά εισάγονται σε ένα αποπολυπλέκτη AWG και στη συνέχεια λαμβάνονται από δύο φωτοηλεκτρικούς δέκτες 11 GHz (U2t Photonics) που είναι συνδεδεμένοι σε έναν παλμογράφο 33GHz (Agilent) που τα απεικονίζει σε πραγματικό χρόνο (Εικόνα 34). Τα ανακτημένα πακέτα δεδομένων καταγράφηκαν από τον παλμογράφο και αξιολογήθηκαν έτσι ώστε να προσδιοριστεί ο ρυθμός σφάλματος Bit του συστήματος (BER) έναντι της οπτικής ισχύος εισόδου στον δέκτη. Για να διευκολυνθεί η μέτρηση BER, χρησιμοποιήθηκε ένας οπτικός ενισχυτής ίνας ερβίου (EDFA) που ακολουθείται από έναν μεταβλητό οπτικό εξασθενητή (VOA) μεταξύ του πομπού και του υποσυστήματος ανίχνευσης.



**Εικόνα 34:** Τα ίχνη που απεικονίζονται στον παλμογράφο πραγματικού χρόνου τα οποία δείχνουν την ταυτόχρονη λήψη των πακέτων σε διαφορετικά μήκη κύματος (μπλέ και κόκκινα ίχνη). Η κυματομορφή που απεικονίζεται με κίτρινο χρώμα δείχνει το ηλεκτρικό σήμα που παράγεται από τον PPG και εγγράφεται στους οπτικούς φορείς από τον ηλεκτρο-οπτικό διαμορφωτή MZM.

Τα πακέτα που λαμβάνονται από τους δύο δέκτες φαίνονται στην Εικόνα 34. Τα πακέτα που απεικονίζονται με μπλε και κόκκινο χρώμα αντιστοιχούν στα πακέτα που είναι εγγεγραμμένα στα μήκη κύματος  $\lambda_1=1535.822\text{nm}$  και  $\lambda_2=1541.75\text{nm}$  αντίστοιχα, μετά από αποπολυπλεξία και ανίχνευση από τον φωτοδέκτη. Για λόγους σύγκρισης, το ηλεκτρικό

σήμα των πακέτων δεδομένων που παράγονται από τον PPG εμφανίζεται στο ίδιο σχήμα με κίτρινο χρώμα στο κάτω μέρος στην Εικόνα 34. Η Εικόνα 34 δείχνει ότι τα δεδομένα του πρώτου πακέτου είναι αποτυπωμένα σωστά στο  $\lambda_1$ , ενώ τα δεδομένα του δεύτερου πακέτου αποτυπώνονται στο  $\lambda_2$ . Θα πρέπει να σημειωθεί ότι οι μικρές αλλοιώσεις που παρατηρούνται στην αρχή και στο τέλος κάθε πακέτου οφείλονται στον περιορισμό των φωτοδεκτών στις χαμηλές συχνότητες (AC-coupled) που χρησιμοποιήθηκαν στο πείραμα. Πιο συγκεκριμένα μερικές από τις αρμονικές χαμηλότερης συχνότητας του σήματος που σχετίζονται με την περιοδικότητα 2,384 kHz των πακέτων που δημιουργήθηκαν, περιορίζονται από τις δυνατότητες του φωτοδέκτη.

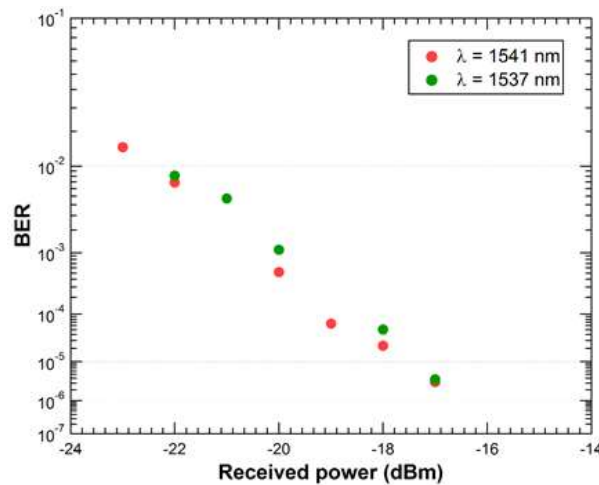
Όπως επιβεβαιώθηκε και από τη μέτρηση BER, αυτές οι μικρές αλλοιώσεις στην αρχή του πακέτου δεν θα έχουν αρνητική επίδραση στη λειτουργία του συστήματος αφού συμπίπτουν με το προοίμιο των πακέτων, το οποίο είναι χρήσιμο στην ανάκτηση ρολογιού στον δέκτη του FPGA. Πιο συγκεκριμένα, αυτές οι αλλοιώσεις μετρήθηκαν έτσι ώστε να μην επηρεάσουν το μήκος του προοιμίου, καθώς σε αυτό περιλαμβάνεται ο χρόνος κλειδώματος του κυκλώματος ρολογιού και ανάκτησης δεδομένων του FPGA (10  $\mu$ s, που αντιστοιχεί σε  $10^5$  bits στα 10 Gb/s). Από την άλλη πλευρά, το τελευταίο κομμάτι του πακέτου συμπίπτει με τον χρόνο παύσης μεταξύ των επόμενων πακέτων και δεν παραμορφώνει τα μεταδιδόμενα δεδομένα (δεν επηρεάζονται τα bit).

Ο ρυθμός σφάλματος bit (BER) φαίνεται στην Εικόνα 35 ως συνάρτηση της ληφθείσας ισχύος για κάθε μήκος κύματος που λαμβάνεται. Ο ρυθμός BER μετρήθηκε μετά την λήψη των δεδομένων από τον παλμογράφο σε πραγματικό χρόνο και την μετέπειτα επεξεργασία τους στο Matlab (offline processing). Η επεξεργασία αυτή προσομοιάζει τη λειτουργία του δέκτη του FPGA που θα προσαρμοστεί σε επόμενα πειράματα. Στα πλαίσια αυτά μελετήθηκε η λειτουργικότητα του δέκτη σχετικά με την ανάκτηση χρονισμού, τον διαχωρισμό του προοιμίου από τα ωφέλιμα δεδομένα, την δειγματοληψία και το κατώφλι για την ανίχνευση συμβόλων και τη μέτρηση BER. Τέλος, η απόδοση του BER εκτιμήθηκε συγκρίνοντας το λαμβανόμενο σήμα με την αρχική ακολουθία bit και μετρώντας τα πραγματικά σφάλματα bit που εμφανίστηκαν κατά τη λήψη. Πρέπει να σημειωθεί ότι η περιορισμένη ικανότητα λήψης δεδομένων του παλμογράφου σε πραγματικό χρόνο (μέγιστο μήκος καταγραφής 20,5 Mpoints ή  $\approx 10,2 \cdot 10^6$  bits στα 40 GSa/s) έθεσε ένα αντίστοιχο όριο στο χαμηλότερο μετρήσιμο BER, το οποίο ήταν επομένως δεσμευμένο στο  $3 \cdot 10^{-7}$  (με ανώτερο διάστημα στατιστικής εμπιστοσύνης 95%). Ωστόσο, οι μετρήσεις BER δεν δείχνουν κατώτατο όριο σφάλματος για επίπεδα ισχύος στον δέκτη ψηλότερα από -17dBm όπου το BER βρίσκεται στην περιοχή μεταξύ 0 και  $3 \cdot 10^{-7}$ .

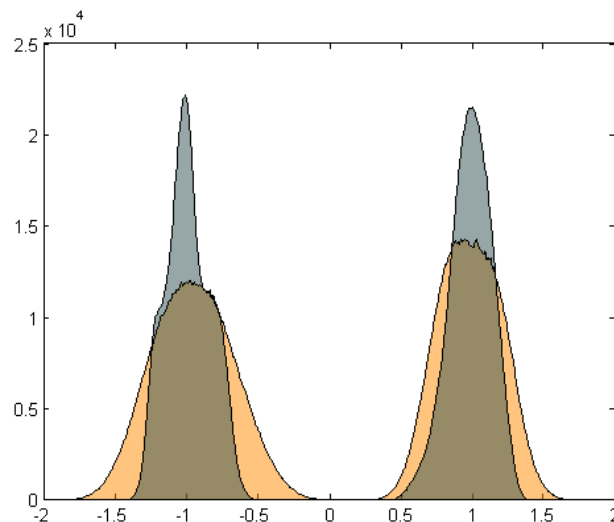
Για να εκτιμηθεί η απόδοση του συστήματος σε αυτήν την περιοχή, χρησιμοποιήθηκαν τα ιστογράμματα συνάρτησης πυκνότητας πιθανότητας (PDF) των λαμβανόμενων σημάτων. Στην Εικόνα 36, εμφανίζονται τα ιστογράμματα PDF των ανιχνευόμενων συμβόλων για  $\lambda_1=1541,75\text{nm}$  σε επίπεδα ισχύος -17dBm (πορτοκαλί ίχνος) και -6 dBm (γκρι ίχνος).

Καθώς αυξάνεται το επίπεδο ισχύος, παρατηρείται σημαντική βελτίωση της ποιότητας του σήματος όσον αφορά τη στατιστική διακύμανση για κάθε κατανομή καθώς και την απόσταση μεταξύ των δύο επιπέδων πλάτους (-1 και 1), υποδεικνύοντας ότι η ποιότητα του σήματος βελτιώνεται περαιτέρω με αυξημένη ισχύ, δηλαδή δεν παρατηρείται κατώτατο όριο σφάλματος.

Σε επόμενη ενότητα θα διερευνηθεί περαιτέρω και θα αξιολογηθεί η απόδοση του BER της πειραματικής διάταξης σε πραγματικό χρόνο, με χρήση FPGA, ώστε τα επίπεδα διαστήματος εμπιστοσύνης να είναι πάνω από  $10^{-13}$ , σύμφωνα με τις τυπικές προδιαγραφές για τα κέντρα δεδομένων όπως ορίζονται στα πρότυπα Ethernet και Infiniband.



**Εικόνα 35:** Απεικόνιση του BER ως συνάρτηση της λαμβανόμενης οπτικής ισχύος για κάθε ένα από τα δύο μήκη κύματος  $\lambda_1=1535.822$  nm (πράσινα σημεία) και  $\lambda_2=1541.75$  nm (κόκκινα σημεία).



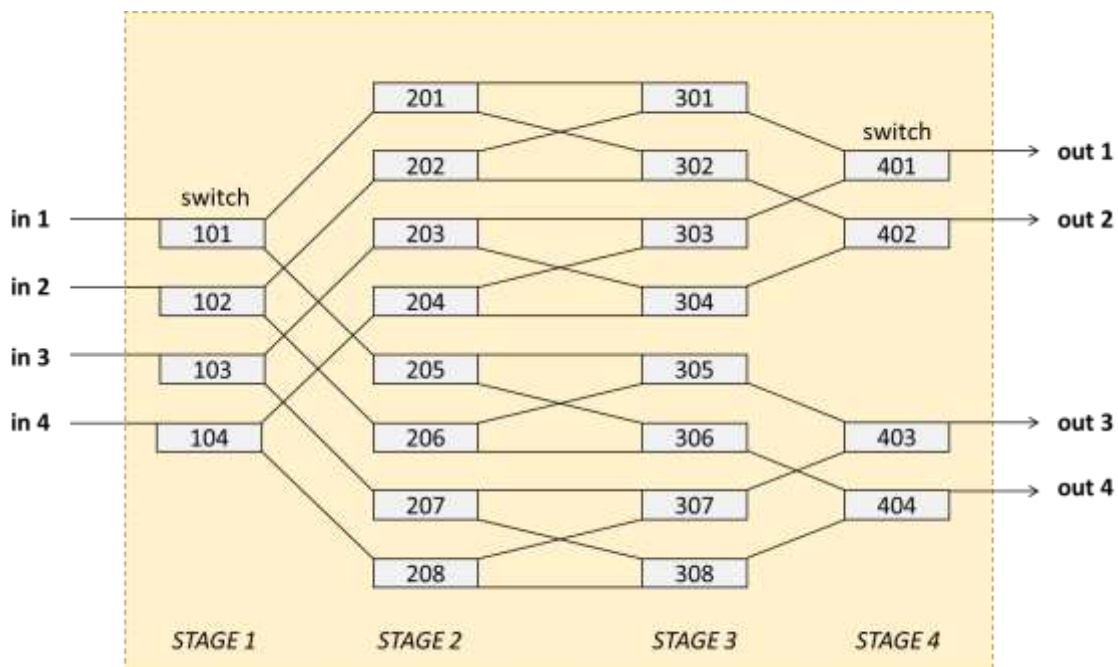


*Εικόνα 36: Ιστογράμματα συνάρτησης πυκνότητας πιθανότητας (Probability Density Function - PDF) των ανιχνευόμενων συμβόλων για  $\lambda_1 = 1541,75 \text{ nm}$  σε λαμβανόμενες οπτικές ισχύεις  $-17 \text{ dBm}$  (πορτοκαλί ίχνος) και  $-6 \text{ dBm}$  (γκρι ίχνος), υποδεικνύοντας βελτιωμένη ποιότητα σήματος για υψηλότερα επίπεδα οπτικής ισχύος.*

### 2.3 Ανάπτυξη γρήγορου $4 \times 4$ οπτικού μεταγωγέα για εξολοκλήρου οπτική μεταγωγή εντός του ToR

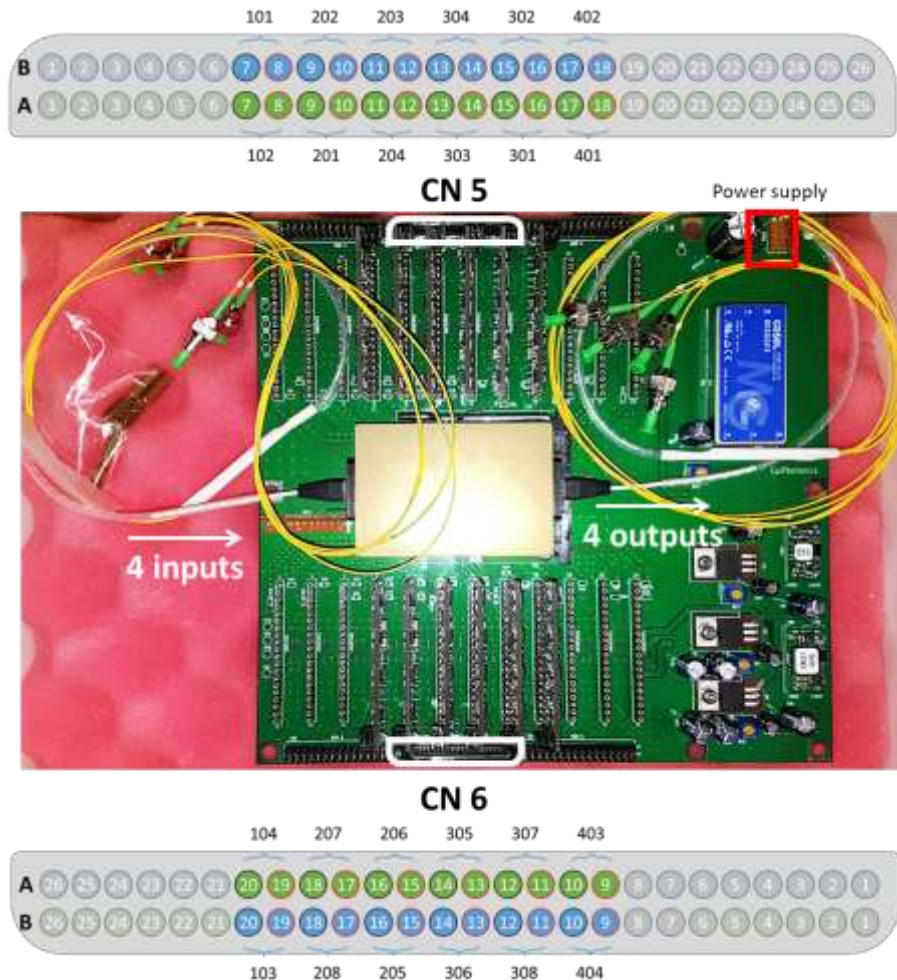
Σε αυτήν την ενότητα γίνεται χαρακτηρισμός ενός γρήγορου οπτικού διακόπτη  $4 \times 4$  που βασίζεται στην τεχνολογία των υλικών  $(\text{Pb,La})(\text{Zr,Ti})\text{O}_3$  [56] και θα αναφέρεται εφεξής ως PLZT διακόπτης. Τα  $(\text{Pb,La})(\text{Zr,Ti})\text{O}_3$  [56] (PLZT) ανήκουν σε μια κατηγορία σιδηροηλεκτρικών κεραμικών υλικών που είναι κατάλληλα σε πολλές ηλεκτροοπτικές εφαρμογές, ενώ παρασκευάζονται από μίκτα οξειδία.

Κατά τη διάρκεια των πειραμάτων χρησιμοποιήθηκε ένας εμπορικά διαθέσιμος γρήγορος  $4 \times 4$  PLZT διακόπτης [57] του οποίου το διάγραμμα κυκλώματος απεικονίζεται στην Εικόνα 37. Ο συγκεκριμένος  $4 \times 4$  PLZT διακόπτης αποτελείται από τέσσερα στάδια μεταγωγής που επιτυγχάνονται μέσω Mach-Zehnder συμβολόμετρων με ηλεκτρο-οπτικές διεπαφές [58]. Μέσω αυτών των σταδίων μπορούν να επιτευχθούν είκοσι-τέσσερις πιθανές καταστάσεις διαμόρφωσης. Έτσι, εφαρμόζοντας τα αντίστοιχα σήματα ελέγχου (Transistor-transistor Logic – TTL) σε κάθε δευτερεύοντα διακόπτη, μπορεί να πραγματοποιηθεί η επιθυμητή οπτική διαδρομή.



*Εικόνα 37: Το εσωτερικό διάγραμμα λειτουργίας του  $4 \times 4$  PLZT διακόπτη, που αποτελείται από 4 στάδια μεταγωγής.*

Η Εικόνα 38 περιλαμβάνει μια φωτογραφία της πλακέτας οπτικού υποσυστήματος 4×4 PLZT όπου η διάταξη οπτικών διακοπών βρίσκεται στο κέντρο της πλακέτας ενώ φαίνονται οι τέσσερις οπτικές ίνες εισόδου και οι τέσσερις οπτικές ίνες εξόδου. Οι ηλεκτρικές διεπαφές που χρησιμοποιούνται για την οδήγηση των τεσσάρων σταδίων «οδήγησης» χωρίζονται σε δύο ομάδες, CN5 και CN6, και βρίσκονται πάνω και κάτω από τον οπτικό διακόπτη αντίστοιχα.

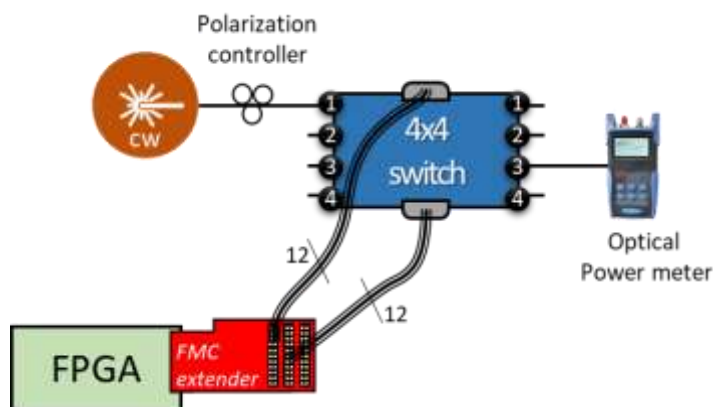


**Εικόνα 38:** Φωτογραφία από τον γρήγορο 4×4 PLZT οπτικό διακόπτη καθώς και το διάγραμμα των ηλεκτρικών συνδέσεων που ελέγχουν τα στάδια μεταγωγής

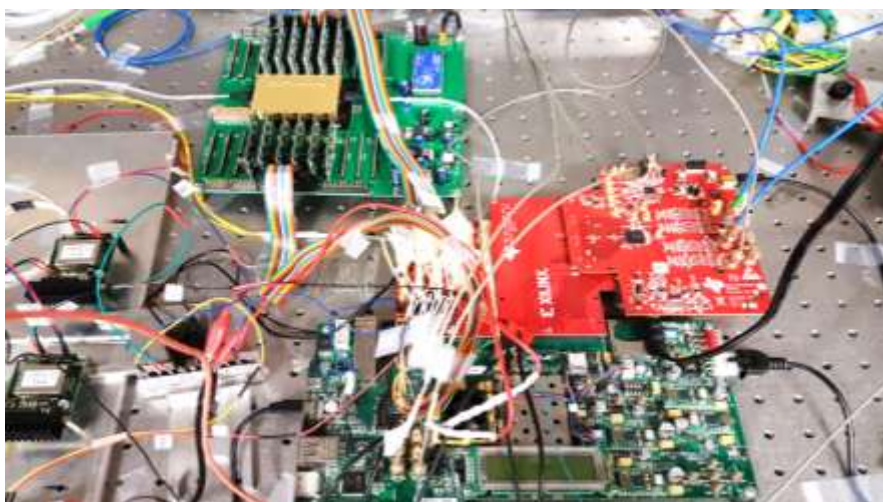
Αρχικά, πραγματοποιήθηκαν τρία διαφορετικά πειράματα προκειμένου να αξιολογηθεί στατικά η απόδοση του γρήγορου 4×4 οπτικού διακόπτη PLZT. Ο στόχος του πρώτου πειράματος ήταν να διερευνηθούν οι απώλειες στην οπτική ισχύ που εισάγονται όταν το φως διέρχεται μέσα από τον οπτικό διακόπτη. Η πειραματική διάταξη απεικονίζεται στην Εικόνα 39. Μια πηγή λέιζερ ακολουθούμενη από έναν ελεγκτή πόλωσης (polarization

controller) εκπέμπει ένα συνεχές οπτικό φέρον ισχύος 10 dBm που εισέρχεται στις θύρες εισόδου του 4×4 οπτικού διακόπτη PLZT.

Για την καταγραφή της οπτικής ισχύος του σήματος που εξέρχεται από κάθε έξοδο του 4×4 οπτικού διακόπτη PLZT, χρησιμοποιήθηκε ένας μετρητής ισχύος που συνδέθηκε σε κάθε μία από τις τέσσερις εξόδους του διακόπτη.



**Εικόνα 39:** Η πειραματική διάταξη που χρησιμοποιήθηκε για τη μέτρηση των απωλειών στην οπτική ισχύ, ύστερα από τη διέλευση του φωτός μέσα από τον διακόπτη.



**Εικόνα 40:** Φωτογραφία που δείχνει την πειραματική διάταξη που χρησιμοποιήθηκε για τον χαρακτηρισμό του γρήγορου 4×4 PLZT οπτικού διακόπτη στο ΕΦΕ

Προκειμένου ο διακόπτης να προγραμματιστεί και να μπορεί να «στρίψει» το φως από τις θύρες εισόδου προς στις θύρες εξόδου χρειάζεται να εφαρμοστούν οι κατάλληλες τάσεις στους ακροδέκτες CN5, CN6 (Εικόνα 38). Για αυτό το λόγο, χρησιμοποιήθηκε ένα FPGA (XILINX KINTEX [59]) με ειδική επέκταση τύπου FPGA Mezzanine Card (FMC) [60], το οποίο είναι ικανό να παράγει ηλεκτρικά σήματα TTL (0-3.3V), που χρησιμεύουν ως σήματα

ελέγχου, όπως φαίνεται στην Εικόνα 40. Αυτά τα σήματα ελέγχου τροφοδοτούν τις αντίστοιχες ηλεκτρικές διεπαφές του και με αυτόν τον τρόπο μπορούν να καθοριστούν δυναμικά οι οπτικές διαδρομές από τις θύρες εισόδου προς τις θύρες εξόδου. Ο Πίνακας 4 περιέχει τα αποτελέσματα για τις απώλειες που εισάγονται κάθε οπτική διαδρομή που προγραμματίζεται. Με βάση αυτά τα αποτελέσματα οι απώλειες κυμαίνονται από 13,5 dB (PATH: είσοδος 4 → έξοδος 2) έως 18,1 dB (PATH: είσοδος 1 → έξοδος 3) με μέση τιμή 14,6 dB. Αξίζει επίσης να σημειωθεί ότι φαίνεται ότι υπάρχει σημαντική εξάρτηση των απωλειών ανάλογα με την πόλωση του εισερχόμενου οπτικού φορέα η οποία κατά μέσο όρο μετρήθηκε στα 2,5dB.

**Πίνακας 4:** Οι απώλειες που εισάγονται σε κάθε οπτική διαδρομή μέσα στον 4×4 οπτικό διακόπτη

Διαδρομή	Απώλειες (dB)	Διαδρομή	Απώλειες (dB)	Διαδρομή	Απώλειες (dB)	Διαδρομή	Απώλειες (dB)
1→1	16.0	2→1	15.5	3→1	16.3	4→1	14.7
1→1	15.9	2→1	15.4	3→1	16.2	4→1	14.7
1→1	15.8	2→1	15.4	3→1	16.2	4→1	14.7
1→1	15.8	2→1	15.4	3→1	16.2	4→1	14.7
1→1	15.8	2→1	15.4	3→1	16.2	4→1	14.7
1→1	15.8	2→1	15.4	3→1	16.2	4→1	14.7
1→1	15.8	2→1	15.4	3→1	16.2	4→1	14.7
1→2	15.8	2→2	14.8	3→2	16.0	4→2	13.6
1→2	15.7	2→2	14.8	3→2	16.0	4→2	13.5
1→2	15.7	2→2	14.9	3→2	16.0	4→2	13.5
1→2	15.7	2→2	14.8	3→2	16.0	4→2	13.5
1→2	15.7	2→2	14.8	3→2	16.0	4→2	13.5
1→2	15.7	2→2	14.8	3→2	16.0	4→2	13.5
1→2	15.7	2→2	14.8	3→2	16.0	4→2	13.5
1→2	15.7	2→2	14.8	3→2	16.0	4→2	13.5
1→3	18.1	2→3	15.5	3→3	16.1	4→3	15.9
1→3	18.0	2→3	15.4	3→3	15.9	4→3	15.8
1→3	18.0	2→3	15.4	3→3	15.9	4→3	15.7
1→3	18.0	2→3	15.4	3→3	15.9	4→3	15.7
1→3	18.0	2→3	15.4	3→3	15.8	4→3	15.7
1→3	18.0	2→3	15.4	3→3	15.8	4→3	15.6
1→4	17.9	2→4	15.4	3→4	16.1	4→4	14.5
1→4	17.9	2→4	15.3	3→4	16.1	4→4	14.4

1→4	17.9	2→4	15.3	3→4	16.1	4→4	14.4
1→4	17.8	2→4	15.3	3→4	16.1	4→4	14.3
1→4	17.8	2→4	15.3	3→4	16.1	4→4	14.4
1→4	17.8	2→4	15.3	3→4	16.1	4→4	14.4

Σε σύγκριση με τις τυπικές προδιαγραφές του διακόπτη, όπως αυτές καταγράφονται στον Πίνακα 5 και δίνονται από τον κατασκευαστή, οι μετρήσεις αποκλίνουν σημαντικά τόσο για τις απώλειες όσο και για την εξάρτησή τους από την πόλωση του εισερχόμενου φωτός. Όπως φαίνεται από τα αποτελέσματα και από την πειραματική διαδικασία, ο διακόπτης 4×4 δεν κατάφερε να επιτύχει τις θεωρητικές προδιαγραφές του προϊόντος. Ως αποτέλεσμα, παρά το γεγονός ότι οι απώλειες που εισάγονται από τη συσκευή είναι υψηλές, ήταν δυνατό να πραγματοποιηθούν περεταίρω πειράματα με την χρησιμοποίηση ενός οπτικού ενισχυτή ίνας ερβίου (EDFA) στην πειραματική διάταξη. Επιπλέον, ένα γενικότερο συμπέρασμα είναι ότι η συγκεκριμένη τεχνολογία δεν έχει φτάσει στο επίπεδο ωριμότητας για πρακτική εφαρμογή.

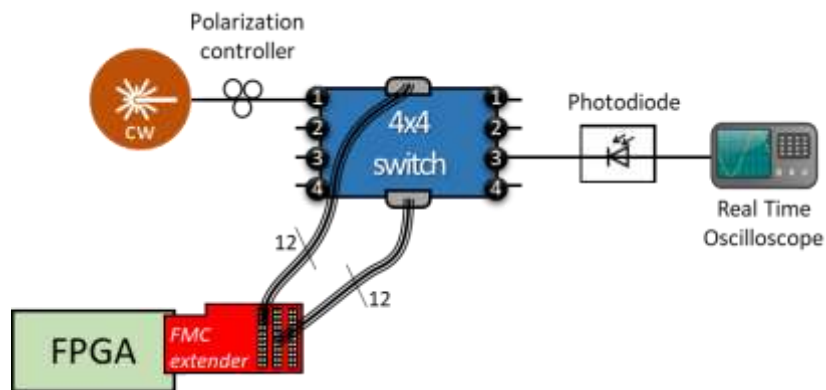
**Πίνακας 5:** Οι τυπικές προδιαγραφές του 4×4 οπτικού διακόπτη PLZT.

Ports	Product #	Major specifications	
4×4	EPS0404S	Architecture	Strictly non-blocking
		Board mounted type	Switch module mounted on driver board
		Rack mounted type	Switch board mounted in 1U-19" case
		Driver control	TTL signals
		Switching time	10 ns typ
		Insertion loss	7.0 dB typ
		Crosstalk	30 dB typ
		PDL	1.0 dB typ

Το δεύτερο πείραμα που περιγράφεται στη συνέχεια αφορά στη μέτρηση του χρόνου μεταγωγής, δηλαδή στον χρόνο που μεσολαβεί μεταξύ δύο διαφορετικών σταθερών καταστάσεων του οπτικού διακόπτη. Η πειραματική διάταξη ήταν παρόμοια με την προηγούμενη και απεικονίζεται στην Εικόνα 41. Μια φωτοδίοδος ακολουθούμενη από έναν παλμογράφο πραγματικού χρόνου συνδέθηκε στην έξοδο του διακόπτη έτσι ώστε να μετρηθεί ο χρόνος μεταγωγής της συσκευής, που είναι το άθροισμα των χρόνων ανόδου και πτώσης της οπτικής ισχύος που καταγράφεται στις θύρες εξόδου.

Στην Εικόνα 42 απεικονίζονται οι χρόνοι ανόδου (α) και πτώσης (β) της οπτικής ισχύος στις θύρες εξόδου και με βάση αυτό μετρήθηκε πειραματικά ο μέσος χρόνος μεταγωγής και ήταν περίπου 33 ns. Επιπλέον, κατά τη διάρκεια του πειράματος έγιναν μετρήσεις του χρόνου μεταγωγής μεταβάλλοντας την οπτική πόλωση στο σήμα εισόδου και το συμπέρασμα είναι ότι δεν παρατηρήθηκε καμία αξιοσημείωτη εξάρτηση του χρόνου μεταγωγής από την πόλωση του εισερχόμενου οπτικού σήματος. Αν και η μετρηθείσα τιμή των 33ns είναι σημαντικά πιο αργή από την αναμενόμενη τιμή σύμφωνα με τις τυπικές προδιαγραφές

(τυπικά 10 ns), το αποτέλεσμα είναι εντός των αποδεκτών τιμών της υπό διερεύνηση αρχιτεκτονικής της διατριβής.



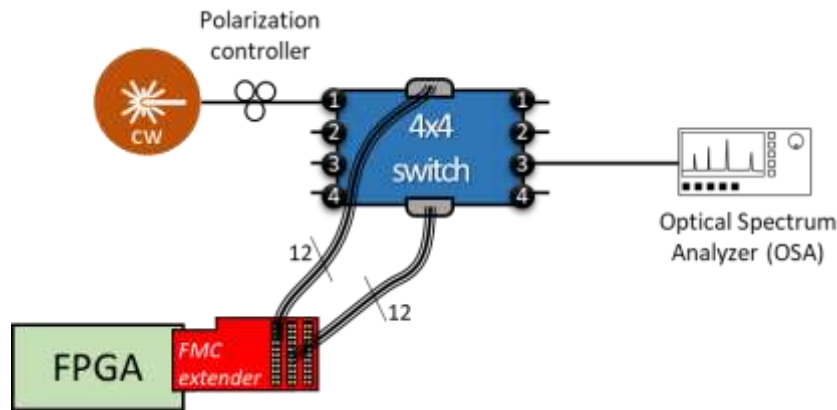
**Εικόνα 41:** Η Πειραματική διάταξη που χρησιμοποιήθηκε για τις μετρήσεις του χρόνου μεταγωγής του οπτικού διακόπτη.



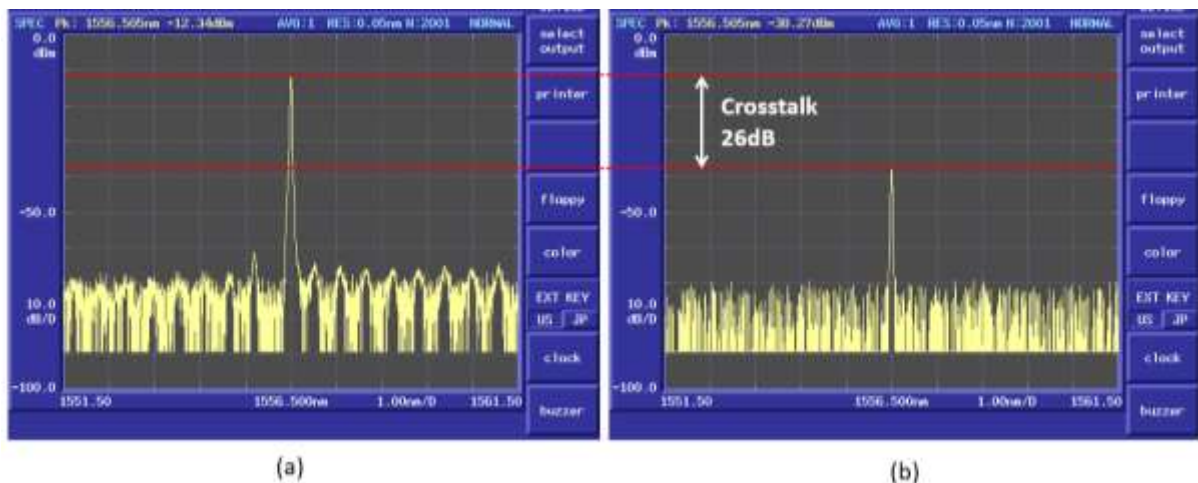
**Εικόνα 42:** Χρόνος μεταγωγής του 4x4 οπτικού διακόπτη PLZT. (α) Χρόνος ανόδου και (β) Χρόνος πτώσης. Οι χρόνοι ανόδου και πτώσης μετρήθηκαν στο 10% και 90% του ρεύματος εξόδου της φωτοδιόδου (μέτρηση 10-90).

Το επόμενο στάδιο της πειραματικής αξιολόγησης του οπτικού διακόπτη αφορά στη μέτρηση της παρεμβολής μεταξύ των καναλιών εξόδου. Για το λόγο αυτό, ο διακόπτης ρυθμίστηκε στατικά σε διάφορες οπτικές διαδρομές και επειδή το αποτέλεσμα ήταν όμοιο σε όλες, παρουσιάζουμε ενδεικτικά τα αποτελέσματα που αφορούν στην οπτική διαδρομή "1→3". Αυτό σημαίνει ότι η οπτική ισχύς που εισέρχεται μέσω της εισόδου 1 κατευθύνεται στην έξοδο 3. Η οπτική ισχύς που μετρήθηκε στις υπόλοιπες θύρες εξόδου ανιχνεύθηκε και

καταγράφηκε μέσω ενός αναλυτή οπτικού φάσματος (Optical Spectrum Analyzer - OSA). Η οπτική ισχύς που μετρήθηκε στην έξοδο 3 απεικονίζεται στην Εικόνα 44 (α) ενώ η οπτική ισχύς που μετρήθηκε στις υπόλοιπες εξόδους στο στην Εικόνα 44 (β). Ο λόγος παρεμβολής μετρήθηκε στα 26 dB, ελαφρώς χειρότερος από την τυπική τιμή (30dB).



*Εικόνα 43: Η πειραματική διάταξη για τη μέτρηση της παρεμβολής (crosstalk)*

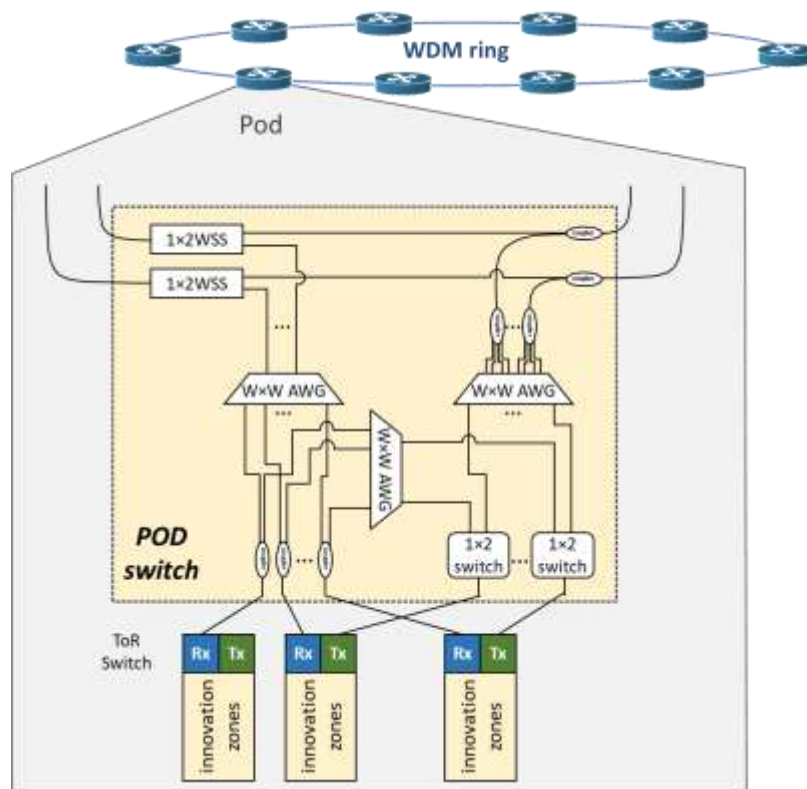


*Εικόνα 44: Μέτρηση οπτικής ισχύος (α) στην έξοδο 3, σε σύγκριση με (β) τις υπόλοιπες εξόδους.*

## 2.4 Ανάπτυξη οπτικών υποσυστημάτων και τεχνικών δρομολόγησης εντός του pod switch

Ο pod switch, όπως έχει αναφερθεί και κατά την περιγραφή της προτεινόμενης αρχιτεκτονικής [61], καθορίζει την δρομολόγηση των πακέτων μεταξύ και εντός του pod και υλοποιείται με δύο τύπους ενεργών οπτοηλεκτρονικών φωτονικών μονάδων. Η μία από αυτές τις μονάδες είναι ένας  $1 \times 2$  οπτικός γρήγορος διακόπτης, ενώ η δεύτερη είναι ο διακόπτης επιλογής μήκους κύματος (WSS). Μια σχηματική εικόνα του διακόπτη POD

δίνεται στην Εικόνα 45. Αν φανταστούμε ότι υπάρχει κίνηση στον κατακόρυφο και στον οριζόντιο άξονα, τότε οι διακόπτες  $1 \times 2$  κατευθύνουν την κίνηση είτε σε άλλο ToR του ίδιου POD (intra-pod) είτε στο WDM δακτύλιο (inter-pod). Από την άλλη οι διακόπτες επιλογής μήκους κύματος υποδέχονται την κίνηση από τον WDM δακτύλιο και διαχωρίζουν τα πακέτα ανά μήκος κύματος ανάλογα με το αν ο προορισμός τους βρίσκεται στο pod που εισέρχονται ή θα πρέπει να προωθηθούν σε κάποιο επόμενο μέσω του WDM δακτυλίου.



*Εικόνα 45: Σχηματική εικόνα του pod switch και διασύνδεση με ToR switches*

#### 2.4.1 Διακόπτης Καθορισμού intra- και inter- pod επικοινωνίας

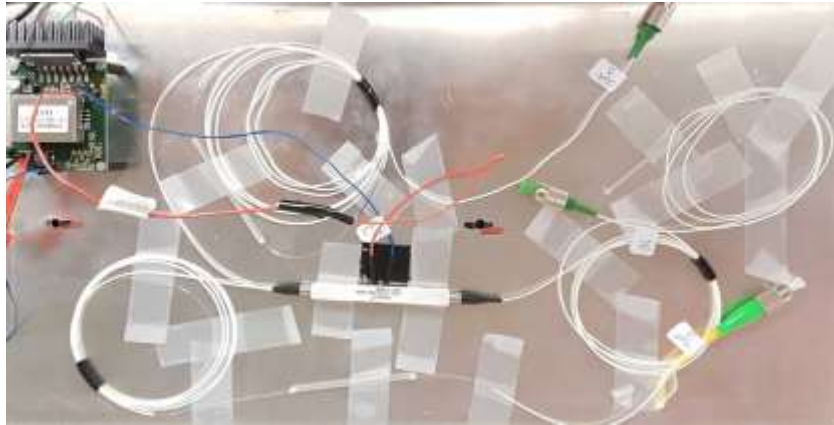
Ο  $1 \times 2$  διακόπτης καθορισμού intra- και inter- pod επικοινωνίας, υλοποιείται μέσω εμπορικά διαθέσιμων οπτικών διακοπών. Πιο συγκεκριμένα στην υλοποίηση που ακολουθείται, χρησιμοποιήθηκαν δύο εμπορικά διαθέσιμοι διακόπτες:

- $2 \times 2$  PLZT διακόπτης βασισμένος στην τεχνολογία  $(\text{Pb}, \text{La})(\text{Zr}, \text{Ti})\text{O}_3$  (PLZT) [58], ο οποίος φαίνεται στην Εικόνα 46.
- $2 \times 2$  οπτικός διακόπτης βασισμένος σε οπτοκεραμική τεχνολογία από την εταιρεία Boston Applied Technologies (BATi) [62], που φαίνεται στην Εικόνα 47.





*Εικόνα 46: Φωτογραφία του 2×2 PLZT διακόπτη της Epi-Photonics*



*Εικόνα 47: Φωτογραφία του 1×2 οπτικού διακόπτη της BATi*

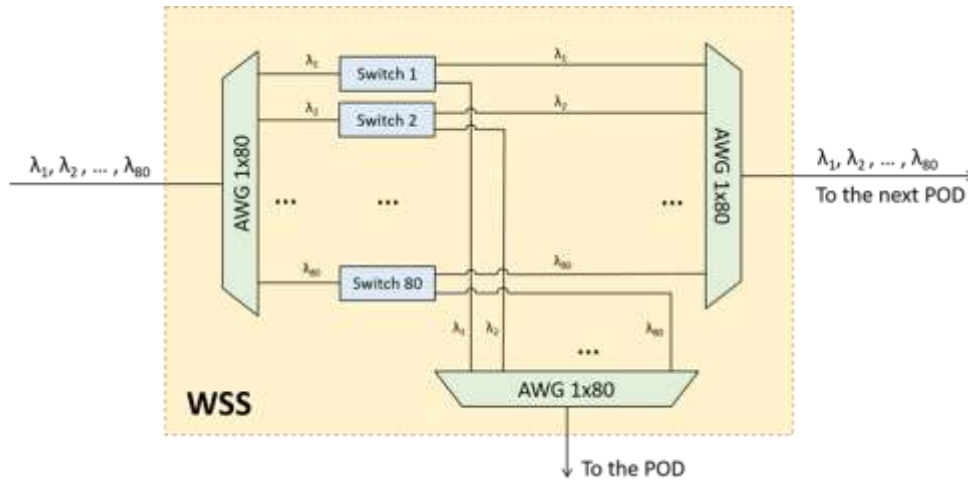
Ο χαρακτηρισμός των παραπάνω συσκευών δεν είναι αντικείμενο της παρούσας διατριβής, απλώς αναφέρονται για λόγους πληρότητας. Αξίζει να σημειωθεί ότι και οι δύο τύποι οπτικών διακοπών πληρούν όλες τις προδιαγραφές της προτεινόμενης αρχιτεκτονικής από άποψη ταχύτητας μεταγωγής, απωλειών ισχύος και τρόπου ελέγχου και διασύνδεσης.

#### **2.4.2 Διακόπτης επιλογής μήκους κύματος (Wavelength Selective Switch)**

Ένα από τα κύρια οπτικά στοιχεία του pod switch είναι ο διακόπτης επιλογής μήκους κύματος (WSS), ο οποίος χρησιμοποιείται για τη διαχείριση της εισερχόμενης κίνησης από τον DWDM δακτύλιο. Στην γενική αρχιτεκτονική για κάθε plane, χρησιμοποιείται ένα διαφορετικό WSS, που εξυπηρετεί όλα τα μήκη κύματος που διαδίδονται μέσω του συγκεκριμένου επιπέδου στο DWDM δακτύλιο.

Η οπτική αρχιτεκτονική WSS που υιοθετήθηκε στην προκειμένη περίπτωση ακολουθεί την προσέγγιση «αποπολυπλεξία, μεταγωγή και αποπολυπλεξία» (demultiplex-switch-

multiplex) που φαίνεται σχηματικά στο σχήμα της Εικόνας 48. Χρησιμοποιούνται τρία AWG 1×80 [63], ένα για αποπολυπλεξία και δύο για πολυπλεξία των οπτικών σημάτων που μεταφέρουν τα πακέτα με βάση το μήκος κύματος. Ενδιάμεσα, χρησιμοποιούνται 80 γρήγοροι διακόπτες 1×2, ένας ανά μήκος κύματος, που προωθούν την κίνηση είτε μέσα στο pod είτε να προωθηθεί στο επόμενο pod σύμφωνα με τις απαιτήσεις του δικτύου.



**Εικόνα 48:** Αρχιτεκτονική της μονάδας διακόπτη επιλογής μήκους κύματος (WSS) με βάση την προσέγγιση «αποπολυπλεξία, μεταγωγή και αποπολυπλεξία».

### 2.4.3 Παθητικός Οπτικός Δρομολογητής (AWGR)

Στο εσωτερικό ενός pod switch χρησιμοποιούνται επίσης παθητικά οπτικά στοιχεία δρομολόγησης που καθορίζουν τόσο την κίνηση εντός όσο και αυτή μεταξύ των pod switches. Πιο συγκεκριμένα δύο παθητικά 80×80 AWGR είναι απαραίτητα εντός κάθε pod switch για κάθε οπτικό plane. Με βάση το μήκος κύματος που είναι εγγεγραμμένο κάθε πακέτο που διέρχεται, η κίνηση δρομολογείται στον αντίστοιχο προορισμό ToR εντός του pod switch ή στον οπτικό δακτύλιο εάν η κίνηση προορίζεται για διαφορετικό pod. Στη συνέχεια, σε όλες οι πειραματικές διατάξεις που παρουσιάζονται χρησιμοποιούνται παθητικοί κυκλικό οπτικοί δρομολογητές (cyclic AWGRs) 8 εισόδων και 8 εξόδων. Πιο συγκεκριμένα, ο κυκλικός τρόπος λειτουργίας τους αποτυπώνεται στον Πίνακα 6.

**Πίνακας 6:** Σχέση καναλιών εισόδου και εξόδου του  $8 \times 8$  cyclic AWGR που χρησιμοποιήθηκε στο εσωτερικό ενός pod switch

nm		Output Ports							
		1	2	3	4	5	6	7	8
Input Ports	1	1550.918	1551.721	1552.524	1546.917	1547.715	1548.515	1549.315	1550.116
	2	1551.721	1552.524	1546.917	1547.715	1548.515	1549.315	1550.116	1550.918
	3	1552.524	1546.917	1547.715	1548.515	1549.315	1550.116	1550.918	1551.721
	4	1546.917	1547.715	1548.515	1549.315	1550.116	1550.918	1551.721	1552.524
	5	1547.715	1548.515	1549.315	1550.116	1550.918	1551.721	1552.524	1546.917
	6	1548.515	1549.315	1550.116	1550.918	1551.721	1552.524	1546.917	1547.715
	7	1549.315	1550.116	1550.918	1551.721	1552.524	1546.917	1547.715	1548.515
	8	1550.116	1550.918	1551.721	1552.524	1546.917	1547.715	1548.515	1549.315

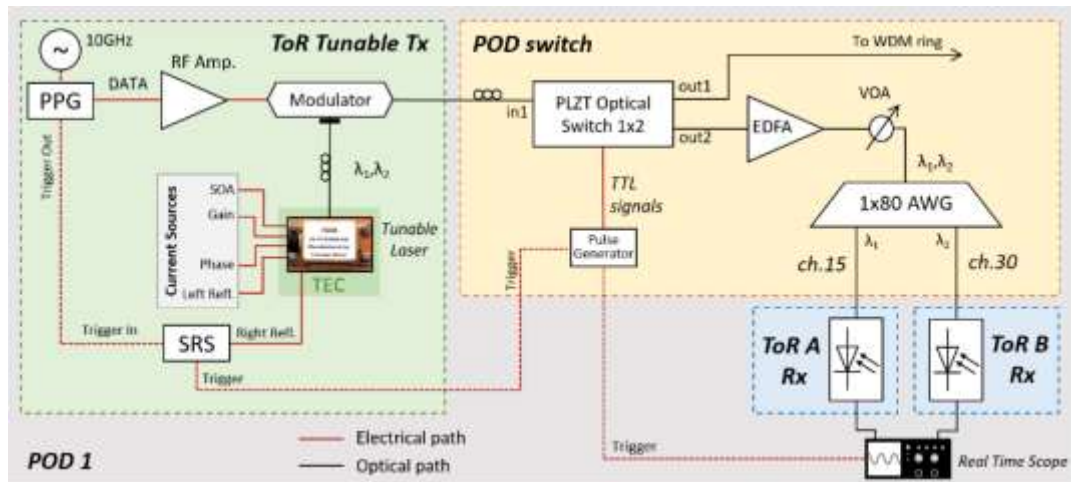
## 2.5 Πειραματική Αξιολόγηση των τεχνικών δρομολόγησης των οπτικών και χαρακτηρισμός των οπτικών διαδρομών

Προκειμένου να επαληθευτεί πειραματικά η επιτυχής λειτουργία της τεχνικής δρομολόγησης των συστημάτων ToR και pod switch της προτεινόμενης αρχιτεκτονικής, πραγματοποιήθηκε μια σειρά πειραμάτων. Στις επόμενες δύο ενότητες παρουσιάζονται οι πειραματικές διαδικασίες προς επιβεβαίωση της ορθής τεχνικής για δρομολόγηση πακέτων από ένα pod στο ίδιο ή σε κάποιο άλλο.

### 2.5.1 Πειραματική αξιολόγηση της δρομολόγησης εντός pod (intra-pod)

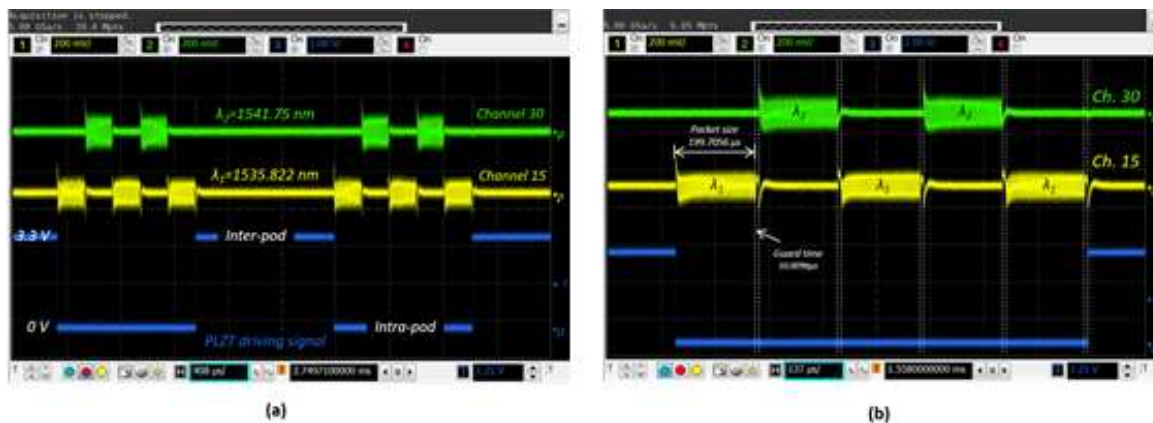
Αρχικά, εφαρμόστηκε το απλοποιημένο σενάριο διασύνδεσης που φαίνεται στην Εικόνα 49 το οποίο περιλαμβάνει τον προσαρμοσμένο πομπό (Ενότητα 2.1) στη μεριά του ToR switch, ο οποίος δημιουργεί πακέτα με προορισμό δύο δέκτες που αντιστοιχούν σε δύο ToR switch (ToR A και ToR B) εντός του ίδιου pod.

Ο πομπός διαμορφώθηκε ώστε να δημιουργεί πακέτα δεδομένων 10 Gb/s σε δύο μήκη κύματος ( $\lambda_1=1535,822$  nm και  $\lambda_2=1541,75$  nm) εναλλάξ. Τα πακέτα που γράφονται στο  $\lambda_1$  έχουν προορισμό το ToR A ενώ από την άλλη τα πακέτα στο  $\lambda_2$  έχουν προορισμό το ToR B. Επιπλέον, στο pod switch ο διακόπτης καθορισμού υλοποιήθηκε με ένα  $1 \times 2$  PLZT οπτικό διακόπτη διασυνδεδεμένο με ένα φίλτρο  $1 \times 80$  AWG που μιμείται τη λειτουργία του  $W \times W$  AWGR της προτεινόμενης αρχιτεκτονικής. Σε αυτό το πείραμα, ο διακόπτης επιλογής μήκους κύματος (WSS) δεν εφαρμόζεται καθώς χειρίζεται τη δρομολόγηση μεταξύ των pod. Στην πλευρά του δέκτη, δύο φωτοδέκτες 10 Gb/s συνδέθηκαν στις αντίστοιχες εξόδους του pod switch και τα φωτορεύματά τους καταγράφηκαν σε παλμογράφο σε πραγματικό χρόνο και αποθηκεύτηκαν για περαιτέρω επεξεργασία.



**Εικόνα 49:** Πειραματική διάταξη για την αξιολόγηση της δρομολόγησης εντός pod

Η Εικόνα 50 (α) απεικονίζει τα πακέτα που μεταδίδονται στα μήκη κύματος  $\lambda_1$  (κίτρινο ίχνος) και  $\lambda_2$  (πράσινο ίχνος) έπειτα από την λήψη τους από τους δέκτες των δύο διαφορετικών ToR μέσα στο ίδιο pod. Το μπλε ίχνος που φαίνεται στην ίδια εικόνα αντιπροσωπεύει το ψηφιακό σήμα ελέγχου TTL για τον 1x2 οπτικό διακόπτη καθορισμού intra- και inter-pod επικοινωνίας. Πιο συγκεκριμένα, όταν η τάση που εφαρμόζεται στον 1x2 οπτικό διακόπτη είναι μηδενική, η οπτική ροή σημάτων κατευθύνεται μέσα στο ίδιο pod, δηλαδή στο out 2 όπως φαίνεται στο σχήμα 49. Αντιθέτως, όταν η ψηφιακή τάση ελέγχου ρυθμιστεί στα 3.3V, η κίνηση οδηγείται έξω από το pod, δηλαδή προς τον οπτικό δακτύλιο WDM.

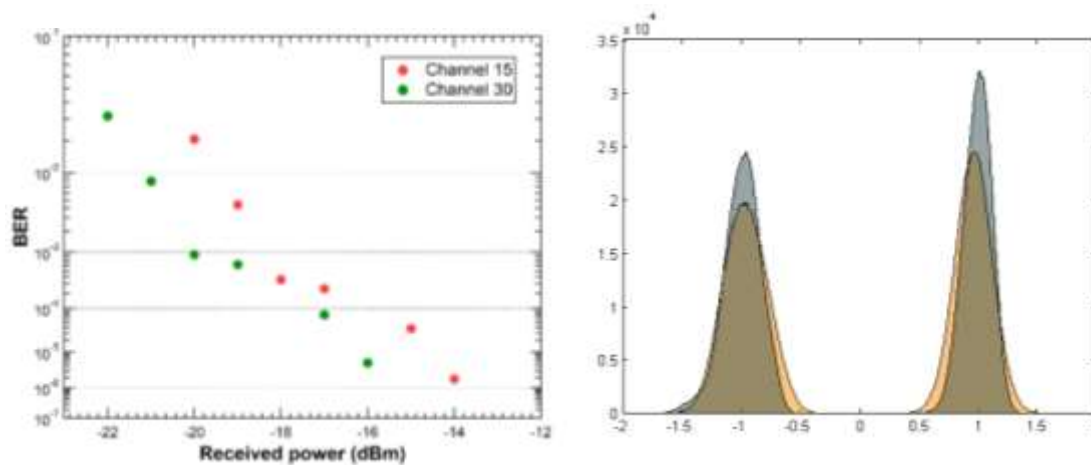


**Εικόνα 50:** (α) Τα πακέτα που εκπέμπονται από τον πομπό στα μήκη κύματος  $\lambda_1$  (κίτρινη κυματομορφή) και  $\lambda_2$  (πράσινη κυματομορφή). Η μπλε κυματομορφή αντιπροσωπεύει το σήμα ελέγχου TTL για τον 1x2 οπτικό διακόπτη καθορισμού. (β) Μεγεθυμένο στιγμιότυπο του (α)

όπου ο χρόνος παύσης ( $\sim 10 \mu s$ ) μεταξύ των πακέτων σημειώνεται με τις διακεκομμένες γραμμές.

Η Εικόνα 50 (β) απεικονίζεται ένα στιγμιότυπο που έχει εστιάσει στα πακέτα που δρομολογούνται εντός του pod, ενώ ο χρόνος παύσης ( $10 \mu s$ ) μεταξύ των πακέτων σημειώνεται με διακεκομμένες γραμμές. Όπως και στις πειραματικές δοκιμές του προσαρμοζόμενου πομπού, η αξιολόγηση συστήματος γίνεται μέσω μέτρησης Bit Error Rate (BER) που πραγματοποιήθηκε μετά τη λήψη δεδομένων σε παλμογράφο και την εκτός σύνδεσης επεξεργασία τους στο Matlab.

Ο ρυθμός εσφαλμένων bit (BER) που απεικονίζεται στην Εικόνα 51 (α) αξιολογήθηκε συγκρίνοντας το σήμα που ανιχνεύτηκε με την αρχική ακολουθία bit και μετρώντας τα πραγματικά σφάλματα bit που εμφανίστηκαν κατά τη λήψη. Αξίζει να σημειωθεί ότι η περιορισμένη ικανότητα λήψης δεδομένων του παλμογράφου έθεσε ένα αντίστοιχο όριο για το χαμηλότερο μετρήσιμο BER, το οποίο καθορίστηκε σε  $3 \cdot 10^{-7}$  (με ανώτερο στατιστικό διάστημα εμπιστοσύνης 95%). Ως πρόσθετη μέτρηση της απόδοσης του συστήματος, διερευνήθηκαν τα ιστογράμματα της συνάρτησης πυκνότητας πιθανότητας των συμβόλων που ανιχνεύθηκαν για διαφορετικά λαμβανόμενα επίπεδα οπτικής ισχύος, όπως φαίνεται στην Εικόνα 51 (β), υποδεικνύοντας ότι δεν υπάρχει κατώτατο όριο σφάλματος.

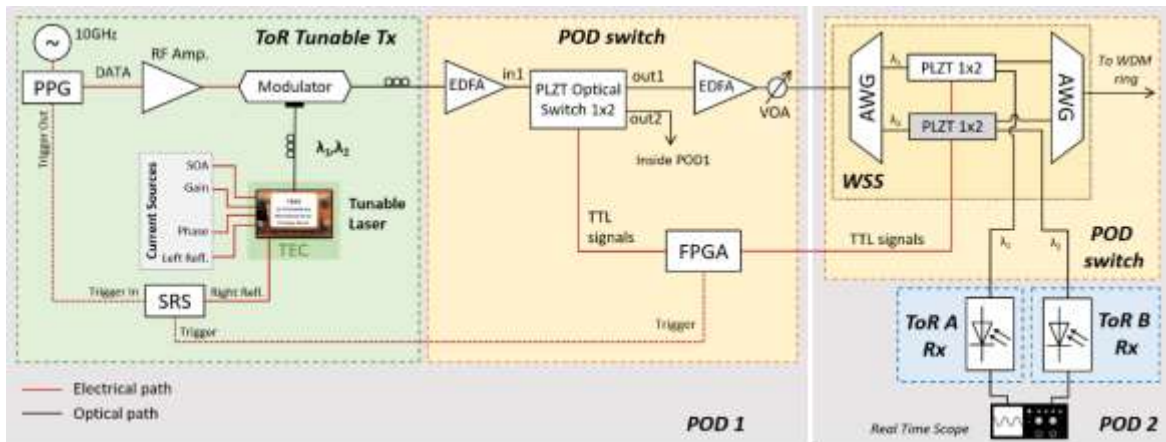


**Εικόνα 51:** α) Μέτρηση BER ως συνάρτηση της λαμβανόμενης οπτικής ισχύος για τα ch.15 και ch.30, β) αντίστοιχα ιστογράμματα της πυκνότητας πιθανότητας (PDF) των ανιχνευθέντων συμβόλων στις λαμβανόμενες οπτικές δυνάμεις  $-14 \text{ dBm}$  (πορτοκαλί ίχνος) και  $-6 \text{ dBm}$  (γκρι ίχνος).

## 2.5.2 Πειραματική απόδειξη της δρομολόγησης εκτός pod (inter-pod)

Με παρόμοιο τρόπο, η δρομολόγηση πακέτων μεταξύ διαφορετικών pod δοκιμάστηκαν μέσω της πειραματικής διάταξης που φαίνεται στην Εικόνα 52. Στην διάταξη αυτή, τρεις

1×2 οπτικοί διακόπτες PLZT είναι απαραίτητοι: ένας που τοποθετείται στην πλευρά του pod 1 και δύο στην πλευρά του pod 2 και χρησιμεύουν για την υλοποίηση του διακόπτη επιλογής μήκους κύματος (Wavelength Selective Switch – WSS).

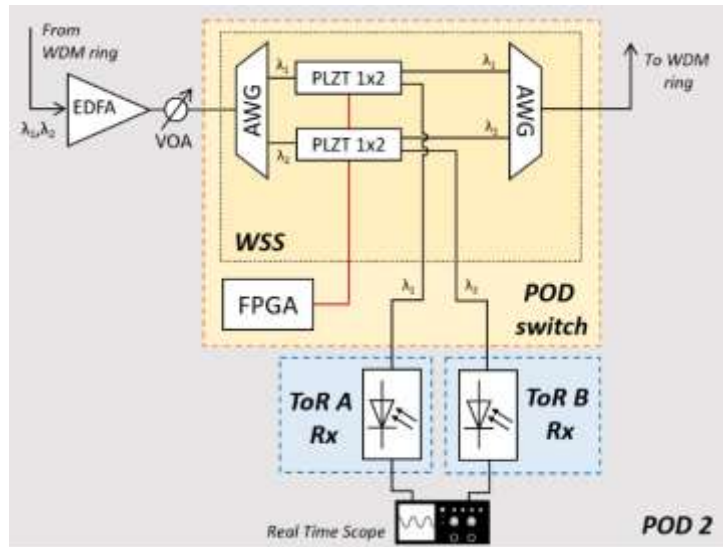


*Εικόνα 52: Πειραματική διάταξη για την αξιολόγηση της δρομολόγησης εκτός pod*

Λόγω μη διαθεσιμότητας όλων των απαιτούμενων στοιχείων τη στιγμή του πειράματος, ήταν απαραίτητο να διαχωριστεί το συγκεκριμένο σενάριο δρομολόγησης σε δύο διαφορετικές φάσεις προκειμένου αφενός να επαληθευτεί η λειτουργικότητα του συστήματος από την σκοπιά του φυσικού στρώματος και αφετέρου η εφαρμογή της επιλεκτικότητας μήκους κύματος:

- **Φάση I (Ενότητα 2.5.2.1):** υλοποίηση και αξιολόγηση της λειτουργίας του διακόπτη επιλογής μήκους κύματος με δύο παράλληλα στοιχεία μεταγωγής.
- **Φάση II (Ενότητα 2.5.2.2):** αξιολόγηση και δοκιμή δύο διαδοχικών στοιχείων οπτικής μεταγωγής.

### 2.5.2.1 ΦΑΣΗ I: Διακόπτης Επιλογής Μήκους Κύματος



*Εικόνα 53: Πειραματική διάταξη για την αξιολόγηση της λειτουργικότητας του WSS στο inter-pod σενάριο.*

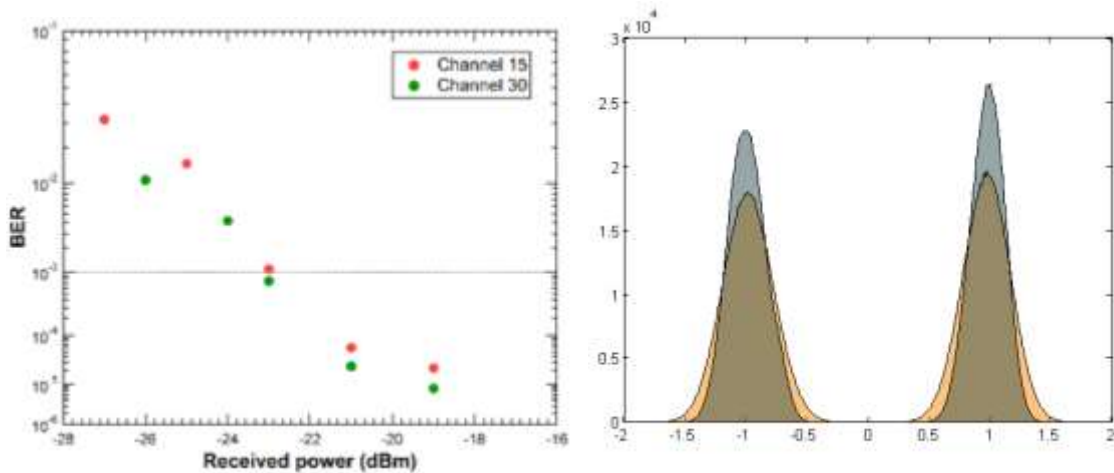
Σε αυτή τη φάση, και οι δύο διαθέσιμοι 1×2 PLZT διακόπτες τοποθετήθηκαν μέσα στο WSS προκειμένου να αξιολογηθεί η λειτουργικότητα της μονάδας. Ως πομπός χρησιμοποιήθηκε ο προσαρμοζόμενος πομπός, όπως έχει περιγραφεί στην ενότητα 2.1. Ο δυναμικός έλεγχος των 1×2 PLZT διακοπών πραγματοποιήθηκε από μια πλακέτα FPGA (Xilinx SPARTAN) που παράγει τα σήματα ελέγχου που φαίνονται στην Εικόνα 54 μαζί με τα πακέτα που αντιστοιχούν σε δύο μήκη κύματος. Η περίοδος του σήματος ελέγχου για το PLZT που αντιστοιχεί στο  $\lambda_1$  (μπλε και πράσινα ίχνη αντίστοιχα) ορίστηκε να είναι 3355,44  $\mu\text{s}$ , και σύμφωνα με αυτή τα πακέτα δρομολογούνται είτε στο εσωτερικό του pod είτε προωθούνται σε διαφορετικό pod μέσω του WDM δακτυλίου.



*Εικόνα 54: α) Τα πράσινα και κίτρινα ίχνη απεικονίζουν τα πακέτα που έχουν προορισμό μέσα στο pod ενώ τα μπλε και κόκκινα ίχνη είναι τα αντίστοιχα σήματα ελέγχου για τους*

διακόπτες PLZT μέσα στο WSS. β) το κίτρινο ίχνος αντιστοιχεί στην οπτική κίνηση μετά την έξοδο του WSS με προορισμό το δακτύλιο WDM.

Όπως και στα προηγούμενα πειράματα, έγινε ψηφιακή επεξεργασία του σήματος (DSP) για την εξομοίωση της λειτουργικότητας του δέκτη και την αξιολόγηση της απόδοσης BER ως συνάρτηση της οπτικής ισχύος και για τα δύο μήκη κύματος, όπως φαίνεται στο σχήμα 55 (α). Τα ιστογράμματα των συναρτήσεων πυκνότητας πιθανότητας (PDF) των λαμβανόμενων συμβόλων που φαίνονται στο σχήμα 55 (β) δείχνουν σημαντική βελτίωση της ποιότητας του σήματος καθώς αυξάνεται το επίπεδο ισχύος.



**Εικόνα 55:** α) BER ως συνάρτηση της λαμβανόμενης οπτικής ισχύος για τα κανάλια 15 ( $\lambda_2$ ) και 30 ( $\lambda_1$ ). β) Ιστογράμματα συνάρτησης πυκνότητας πιθανότητας (PDF) των bits που ανιχνεύθηκαν για λαμβανόμενη οπτική ισχύ  $-19$  dBm (πορτοκαλί ίχνος) και  $-12$  dBm (γκρι ίχνος), υποδεικνύοντας βελτιωμένη ποιότητα σήματος για υψηλότερα επίπεδα οπτικής ισχύος.

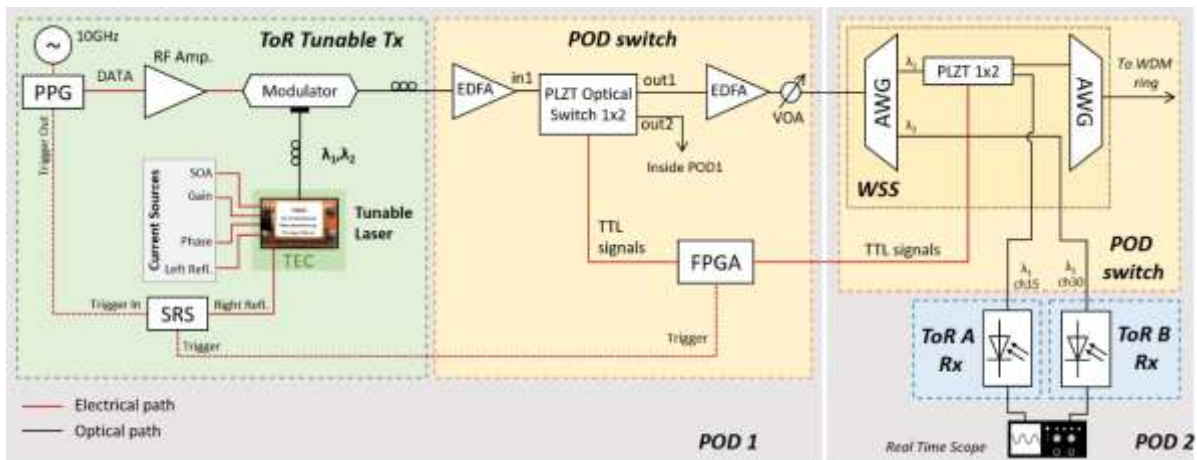
### 2.5.2.2 ΦΑΣΗ II: Αξιολόγηση δύο διαδοχικών στοιχείων οπτικής μεταγωγής.

Η Εικόνα 56 παρουσιάζει την πειραματική διάταξη που χρησιμοποιήθηκε για τη δεύτερη φάση της αξιολόγησης και που αφορά δύο διαδοχικά στοιχεία οπτικής μεταγωγής. Πιο συγκεκριμένα, ένας  $1 \times 2$  οπτικός PLZT διακόπτης χρησιμοποιήθηκε ως διακόπτης καθορισμού intra- και inter-rod δρομολόγησης ως μέρος του rod 1, ενώ το υπόλοιπο της πειραματικής ρύθμισης παρέμεινε το ίδιο με αυτό που περιγράφηκε στην προηγούμενη υποενότητα.

Στην Εικόνα 57 απεικονίζονται τα σήματα ελέγχου για κάθε ένα από τους δύο οπτικούς διακόπτες PLZT και τα μεταδιδόμενα πακέτα στα  $\lambda_1$  και  $\lambda_2$ . Πιο συγκεκριμένα, η μπλε κυματομορφή δείχνει το σήμα ελέγχου για τον διακόπτη που βρίσκεται στο rod 1 (rod πομπού). Σε κάθε περίοδο, 6 πακέτα (τρία σε κάθε μήκος κύματος) δρομολογούνται προς

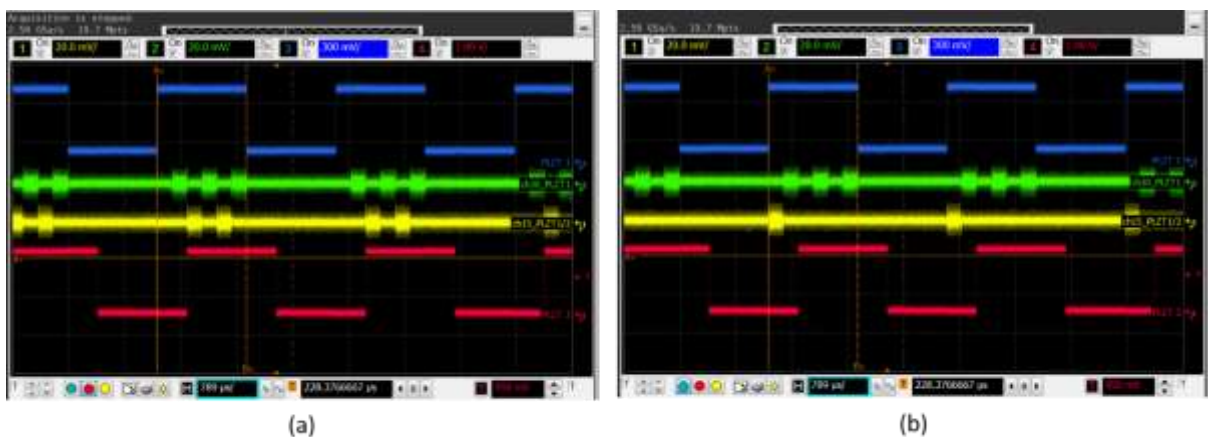


το rod 2 (μέσω της εξόδου 1 του διακόπτη καθορισμού) ενώ τα υπόλοιπα 6 ανακατευθύνονται μέσα στο rod 1 (μέσω της εξόδου 2 του διακόπτη καθορισμού).



**Εικόνα 56:** Πειραματική διάταξη για την αξιολόγηση και δοκιμή δύο διαδοχικών στοιχείων οπτικής μεταγωγής: ένας διακόπτης PLZT στο rod 1 και ένας διακόπτης ως μέρος του WSS στο rod 2.

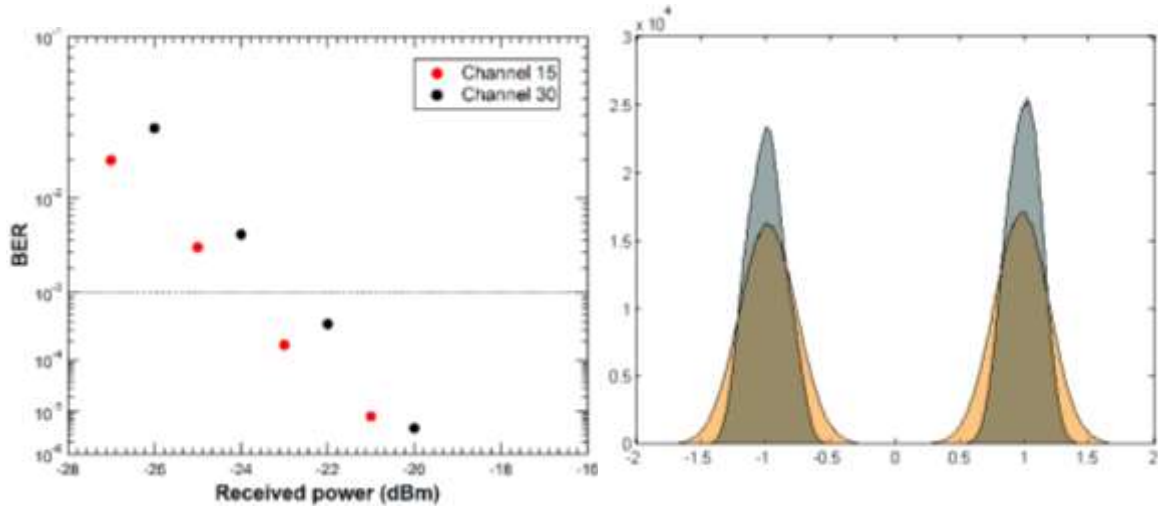
Αντίστοιχα, το σήμα ελέγχου του διακόπτη επιλογής μήκους κύματος που βρίσκεται μέσα στη μονάδα του WSS του rod 2 (κόκκινη κυματομορφή) ελέγχει τον διακόπτη έτσι ώστε για κάθε περίοδο δύο από τα τρία πακέτα στο  $\lambda_1$  να δρομολογούνται στο rod 2 (Εικόνα 57 - α), ενώ το εναπομείναν πακέτο (Εικόνα 57 - β) συνεχίζει μέσω του οπτικού δακτυλίου στο επόμενο rod. Τέλος, το πράσινο ίχνος απεικονίζει τα πακέτα στο  $\lambda_2$  τα οποία διαδίδονται όλα προς τον δακτύλιο WDM και δεν πέφτουν ποτέ στο rod 2.



**Εικόνα 57:** Το μπλε ίχνος δείχνει το σήμα ελέγχου για τον διακόπτη καθορισμού στο rod 1 και το κόκκινο αντιπροσωπεύει το σήμα οδήγησης για τον διακόπτη μέσα στη μονάδα του WSS του rod 2. Η πράσινη κυματομορφή απεικονίζει τα πακέτα που έχουν εγγραφεί στο  $\lambda_2$

(κανάλι 30). (α) Δύο κίτρινα πακέτα στο  $\lambda_1$  (κανάλι 15) προορίζονται για το *rod 2*. (β) Ένα κίτρινο πακέτο στο  $\lambda_1$  (κανάλι 15) προωθείται στον οπτικό δακτύλιο WDM.

Χρησιμοποιήθηκε και πάλι η επεξεργασία DSP εκτός σύνδεσης για την εξομοίωση της λειτουργικότητας του δέκτη και την αξιολόγηση του BER ως συνάρτηση της οπτικής ισχύος του δέκτη και για τα δύο μήκη κύματος, όπως φαίνεται στην Εικόνα 58.



**Εικόνα 58:** α) BER ως συνάρτηση της λαμβανόμενης οπτικής ισχύος για τα κανάλια 15 ( $\lambda_1$ ) και 30 ( $\lambda_2$ ). β) Ιστογράμματα συνάρτησης πυκνότητας πιθανότητας (PDF) των bits που ανιχνεύθηκαν σε λαμβανόμενες οπτικές ισχύς -21 dBm (πορτοκαλί ίχνος) και -14 dBm (γκρι ίχνος), υποδεικνύοντας βελτιωμένη ποιότητα σήματος για υψηλότερα λαμβανόμενα επίπεδα οπτικής ισχύος

## ΚΕΦΑΛΑΙΟ 3

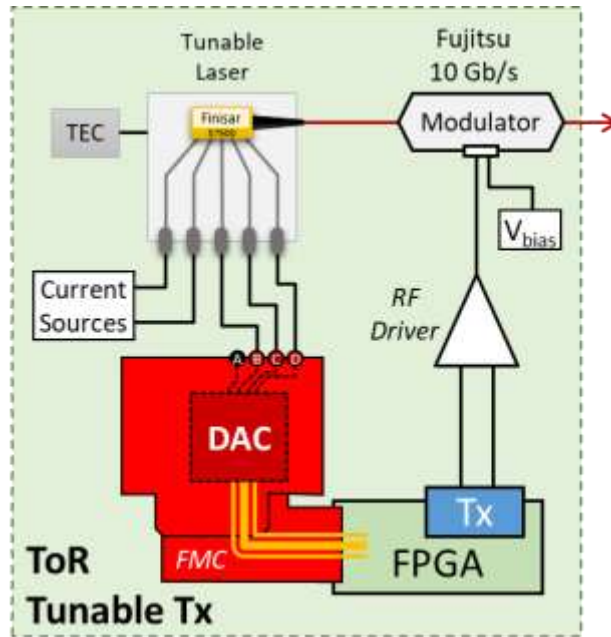
# Ενοποίηση των Οπτικών Υποσυστημάτων και Αξιολόγηση Τεχνικών Δρομολόγησης

Στο κεφάλαιο αυτό αναλύονται και περιγράφονται οι διαδικασίες ενοποίησης και ενσωμάτωσης των επιμέρους οπτικών υποσυστημάτων που αναφέρθηκαν στο ΚΕΦΑΛΑΙΟ 2 και αποτελούν το ToR switch. Επιπλέον, γίνεται εκτενής αναφορά και στον τρόπο διασύνδεσης μεταξύ ToR και pod switch, καθώς και πειραματική αξιολόγηση όλων των πιθανών καναλιών επικοινωνίας.

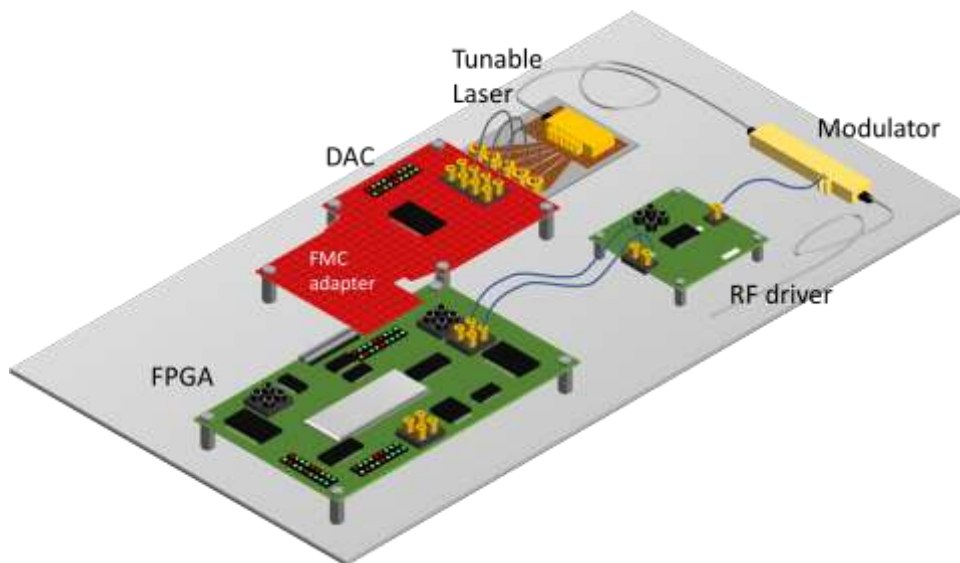
### 3.1 Ενσωμάτωση Προσαρμοζόμενου Πομπού με μετατροπέα ψηφιακής σε αναλογική έξοδο μέσω FPGA

Η Εικόνα 59 δείχνει το διάγραμμα του προσαρμοζόμενου πομπού στο ToR switch που έχει τη δυνατότητα να εκπέμπει φως στο επιθυμητό μήκος κύματος σε όλη τη C μπάντα με ταυτόχρονη διαμόρφωση πακέτων σε ρυθμό 10 Gb/s. Αντίστοιχα, η Εικόνα 60 δείχνει μια 3D απεικόνιση του πομπού στο ToR όπως αυτή σχεδιάστηκε για τον τελικό πρωτότυπο. Πιο αναλυτικά, ο πομπός αποτελείται από:

- Προσαρμοζόμενο λέιζερ (Tunable Laser) [46]
- Ελεγκτή θερμοκρασίας για το λέιζερ (Temperature Controller TEC)
- 10 Gb/s Mach-Zehnder Modulator Οπτικό Διαμορφωτή (MZM)
- Ενισχυτή ραδιοσυχνότητας 10GHz με διαφορική έξοδο (RF driver)
- Μετατροπέα ψηφιακού σε αναλογικό σήμα (Digital-to-analog Converter DAC) [55]
- ToR FPGA: Συστοιχία επιτόπια προγραμματιζόμενων πυλών (Field Programmable Gate Array)
- Προσαρμογέα FPGA Mezzanine Card (FMC) για τη διασύνδεση του DAC στο FPGA.



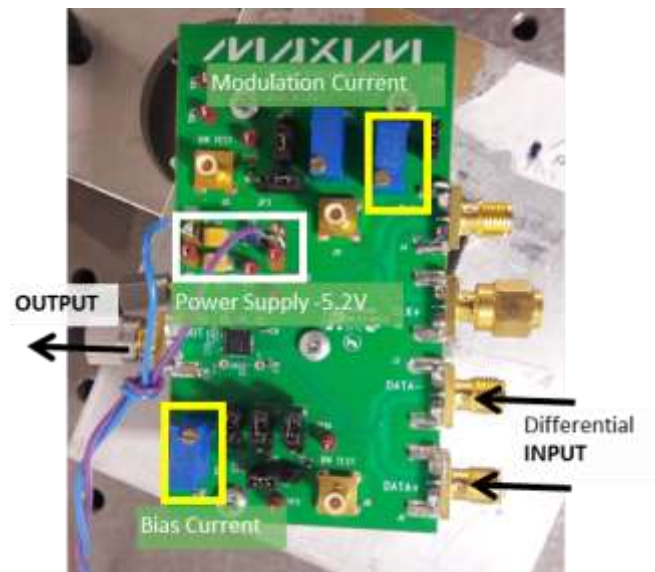
*Εικόνα 59: Το υποσύστημα του προσαρμοζόμενου πομπού στον ToR switch*



*Εικόνα 60: 3D απεικόνιση του προσαρμοζόμενου πομπού όπως σχεδιάστηκε.*

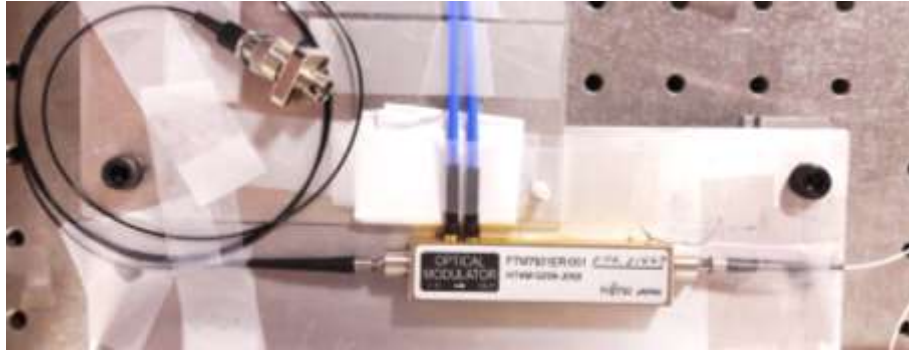
Σε αυτήν την ενότητα επιχειρείται η περιγραφή της ενσωμάτωσης και διασύνδεσης των διαφορετικών μερών που παρατίθενται στην προηγούμενη λίστα και συνθέτουν το ToR switch. Πιο συγκεκριμένα, η παραγωγή των πακέτων προς διαμόρφωση/μετάδοση ως ψηφιακά δυαδικά δεδομένα στα 10 Gb/s (NRZ-10Gb/s) γίνεται από το FPGA. Εν συνεχεία, αυτά τα πακέτα μεταδίδονται ως ηλεκτρικά διαφορικά σήματα μέσω των πομποδεκτών της πλακέτας FPGA. Πριν από την είσοδο στον ηλεκτρο-οπτικό διαμορφωτή, τα σήματα

τροφοδοτούνται στον ενισχυτή ραδιοσυχνότητας 10 Gbps (RF Maxim 3941 - Εικόνα 61) με διαφορική είσοδο (differential input) και μια έξοδο (single ended) προκειμένου να ενισχυθούν μέχρι την απαιτούμενη ταλάντευση τάσης  $V_{\pi}$  που απαιτείται στον διαμορφωτή. Αυτός ο ενισχυτής οδήγησης (RF driver) τροφοδοτείται με  $-5,2\text{ V}$  και όριο ρεύματος στα  $300\text{ mA}$ , αποδίδοντας μια έξοδο που κυμαίνεται από  $1$  έως  $3\text{ V}_{pp}$ . Προκειμένου να βελτιστοποιηθεί το ηλεκτρικό σήμα πριν από την είσοδο στον διαμορφωτή, χρησιμοποιήθηκαν τα ενσωματωμένα στην πλακέτα του RF driver ποτενσιόμετρα για τη σωστή ρύθμιση των ρευμάτων πόλωσης και διαμόρφωσης.



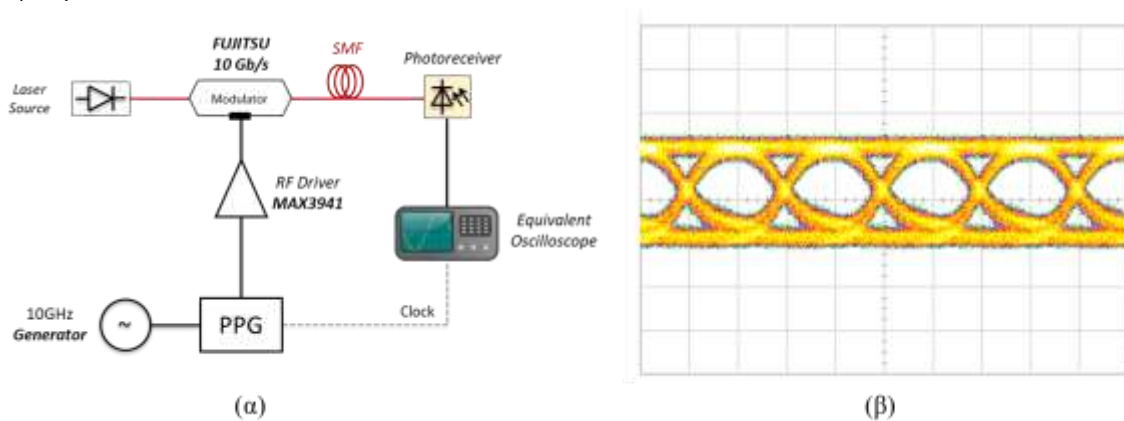
*Εικόνα 61: Ο 10Gb/s ενισχυτής RF που χρησιμοποιείται στον προσαρμοζόμενο πομπό*

Το ενισχυμένο ηλεκτρικό σήμα με τα δεδομένα των πακέτων τροφοδοτεί την αντίστοιχη RF είσοδο ενός Mach-Zehnder διαμορφωτή (MZM) – (Fujitsu 10 Gb/s, Εικόνα 62) ενώ μια ξεχωριστή τάση πόλωσης DC (direct current) τροφοδοτεί τη δεύτερη είσοδο του διαμορφωτή για να διασφαλιστεί ότι επιτυγχάνεται το βέλτιστο σημείο πόλωσης.



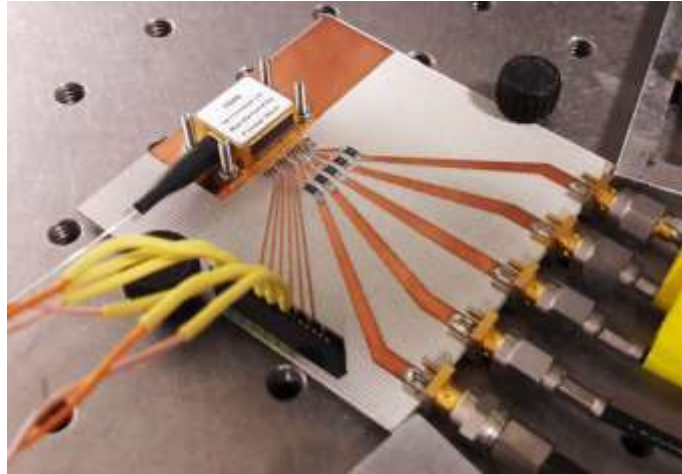
**Εικόνα 62:** 10 Gb/s Mach-Zehnder Modulator Οπτικός Διαμορφωτής (MZM)

Σημειώνεται ότι το σύστημα του ενισχυτή με τον οπτικό διαμορφωτή έχει αξιολογηθεί πειραματικά μέσω της διάταξης που φαίνεται στην Εικόνα 63(α). Πιο συγκεκριμένα, το σύστημα αυτό δοκιμάστηκε σε ρυθμό μετάδοσης 10Gb/s και όπως φαίνεται από το διάγραμμα ματιού (eye diagram) της Εικόνας 63(β) η μετάδοση έγινε επιτυχώς χωρίς σφάλματα.

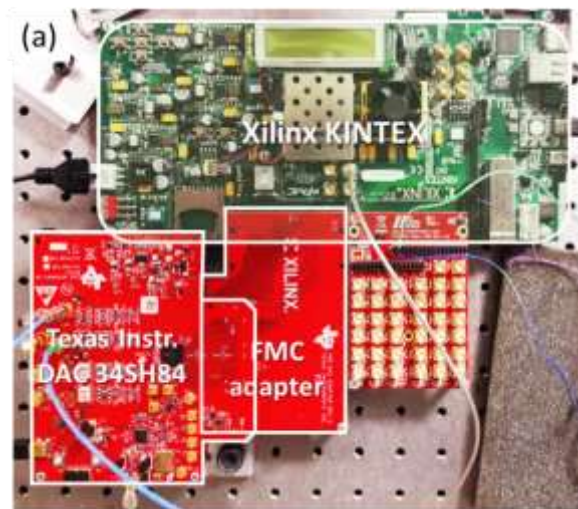


**Εικόνα 63:** (α) Πειραματική διάταξη αξιολόγησης του συστήματος 10Gb/s RF driver – Modulator και (β) Το διάγραμμα ματιού στα 10Gb/s που αποδεικνύει την μετάδοση χωρίς σφάλματα

Ο οπτικός φορέας στο επιθυμητό μήκος κύματος παράγεται από ένα προσαρμοζόμενο λέιζερ (Finisar S7500, Εικόνα 64), το οποίο έχει αναλυθεί και χαρακτηριστεί λεπτομερώς στην ενότητα 2.1. Στο σημείο αυτό, υπενθυμίζεται ότι το προσαρμοζόμενο λέιζερ μπορεί να εκπέμπει στη C-band στο επιθυμητό μήκος κύματος ανάλογα με τα ρεύματα που ελέγχουν τους ανακλαστήρες και τη φάση του, ενώ αντιθέτως τα ρεύματα στους τομείς ενίσχυσης του λέιζερ (SOA και GAIN) μπορούν να παραμένουν σταθερά κατά τη λειτουργία του και γι' αυτό το λόγο παρέχονται από μια σταθερή πηγή ρεύματος.



*Εικόνα 64: Φωτογραφία από το προσαρμοζόμενο λέιζερ Finisar S7500*



*Εικόνα 65: Το σύστημα FPGA/DAC για τον έλεγχο του προσαρμοζόμενου λέιζερ, τον συγχρονισμό των συστημάτων και την παραγωγή των πακέτων της αρχιτεκτονικής*

Προκειμένου να συντονιστεί το λέιζερ στο επιθυμητό μήκος κύματος σύμφωνα με τις ανάγκες του δικτύου, πρέπει να μεταβάλλονται με δυναμικό και σύγχρονο τρόπο τα ρεύματα της φάσης και των ανακλαστήρων στην κοιλότητα του λέιζερ. Για αυτό το λόγο χρησιμοποιείται έναν ψηφιακός σε αναλογικό μετατροπέα (DAC) της Texas Instruments (TI) (DAC34SH84) που παρέχει ακρίβεια 16bit. Μέσω του DAC μια ψηφιακή τιμή μετατρέπεται σε στάθμες αναλογικής τάσης στις εξόδους του. Αυτό με τη σειρά του είναι ταυτόσημο με την παραγωγή διαφορετικών τιμών ρευμάτων, που χρειάζονται για να ελεγχθεί το λέιζερ. Ο μετατροπέας DAC ελέγχεται δυναμικά από το FPGA και η λειτουργία του περιγράφεται στην υποενότητα 3.2.

Τα οπτικά σήματα που εξέρχονται από τον διαμορφωτή οδηγούνται προς στο *pod switch* και στη συνέχεια κατευθύνονται είτε προς διαφορετικό ToR εντός του ίδιου pod είτε προς και διαφορετικό προορισμό pod μέσω του οπτικού δακτυλίου. Το υποσύστημα του προσαρμοζόμενου πομπού, αξιολογήθηκε πειραματικά προκειμένου να επικυρωθεί η λειτουργικότητα του συστήματος καθώς και η επιτυχής ενσωμάτωση των υπόλοιπων τμημάτων του διακόπτη ToR. Μια πιο λεπτομερής περιγραφή αυτής της αξιολόγησης ακολουθεί στην επόμενη ενότητα.

### 3.2 Δυναμικός Έλεγχος Οπτικών Υποσυστημάτων

Το ToR switch όπως έχει αναφερθεί αποτελείται από επιμέρους μέρη και φυσικά απαιτείται πλήρης συγχρονισμός μεταξύ τους για την εύρυθμη λειτουργία του. Αυτό φυσικά, σημαίνει ότι όλα τα μέρη πρέπει να μπορούν να ελέγχονται με δυναμικό και σύγχρονο τρόπο.

Το κύριο οπτικό στοιχείο για το οποίο απαιτείται δυναμικός έλεγχος είναι το προσαρμοζόμενο λέιζερ, που αναλύθηκε στην ενότητα 2.1. Δεδομένου ότι τα εκπεμπόμενα οπτικά πακέτα πρέπει να πληρούν τις απαιτήσεις και να ακολουθούν τον προγραμματισμό του δικτύου της προτεινόμενης αρχιτεκτονικής, το λέιζερ πρέπει να συντονίζεται σε κάθε χρονοθυρίδα (time-slot) έτσι ώστε τα πακέτα να εγγράφονται στο μήκος κύματος που αντιστοιχεί στον ToR switch του προορισμού τους. Υπενθυμίζεται ότι στην προτεινόμενη αρχιτεκτονική κάθε time-slot διαρκεί 200 μs, ενώ ο χρόνος παύσης/ρύθμισης είναι 10 μs.

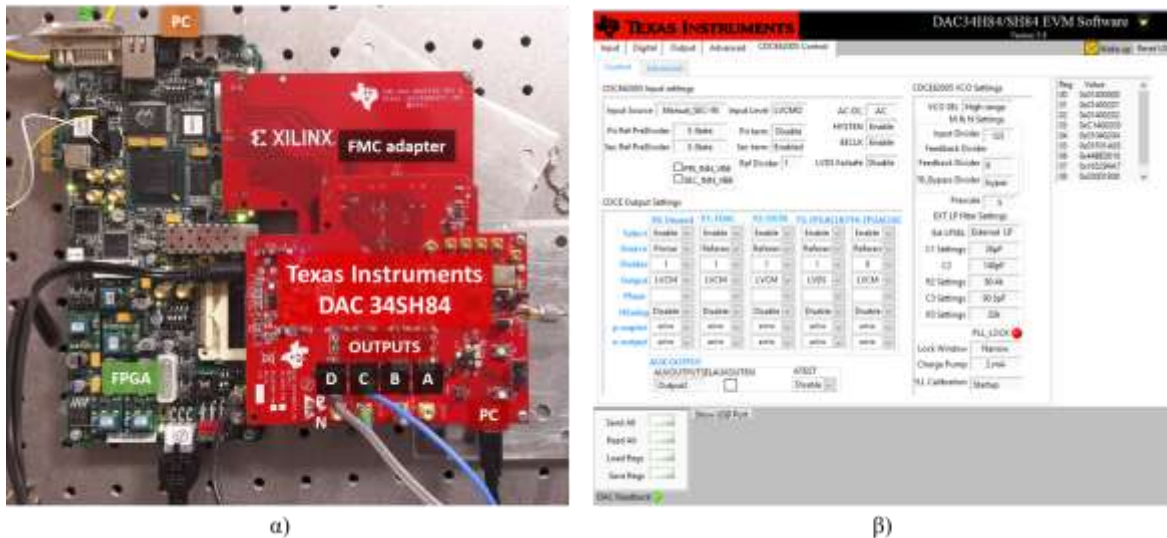
Για να επιτευχθεί λοιπόν ο ανά slot συντονισμός του λέιζερ χρησιμοποιείται ένας μετατροπέας ψηφιακού σε αναλογικό (Digital-to-Analog Converter DAC), όπως περιγράφεται στην ενότητα 3.1 παρέχοντας με δυναμικό τρόπο τα ρεύματα στους τομείς της φάσης και των ανακλαστήρων του προσαρμοζόμενου λέιζερ. Ο DAC ενσωματώνεται πάνω στο FPGA μέσω ενός FMC προσαρμογέα προκειμένου να λάβει τα απαραίτητα ψηφιακά σήματα που αντιστοιχούν στα αναλογικά ρεύματα που θα τροφοδοτηθούν στο συντονιζόμενο λέιζερ.

Η συγκεκριμένη πλακέτα (DAC 34SH84), που φαίνεται στην Εικόνα 66 (α), επιλέχθηκε αφού πληροί όλες τις απαραίτητες απαιτήσεις:

- Διαθέτει τέσσερα διαφορετικά κανάλια που λειτουργούν έως και 1,5 GSa/s, που ξεπερνά κατά πολύ την απαίτηση των 210 μs.
- Παρουσιάζει χρόνο εναλλαγής των σταθμών εξόδου της τάξης των 10 ns, πολύ χαμηλότερο από την απαίτηση των 50 ns στον χρόνο αλλαγής.
- Προσφέρει εύρος ρεύματος εξόδου από 0,2 έως 30 mA, που ταιριάζει καλά στο εύρος των ρευμάτων που απαιτούνται για τον δυναμικό συντονισμό του προσαρμοζόμενου λέιζερ και στα 80 μήκη κύματος της C μπάντας σύμφωνα με το πρότυπο των 50 GHz της ITU.



- Ελέγχεται απευθείας από πλακέτα FPGA μέσω της υποδοχής FMC, διευκολύνοντας τη συγχρονισμένη λειτουργία με τον «ενορχηστρωτή» της προτεινόμενης αρχιτεκτονικής δικτύου και το συνολικό συγχρονισμό του συστήματος.
- Διαθέτει ειδικό λογισμικό με κατάλληλο γραφικό περιβάλλον (Graphical User Interface – GUI) για εύκολη και δυναμικό έλεγχο, επιτρέποντας την άμεση βελτιστοποίηση των παραμέτρων λειτουργίας. Το γραφικό περιβάλλον φαίνεται και στην Εικόνα 66 (β).

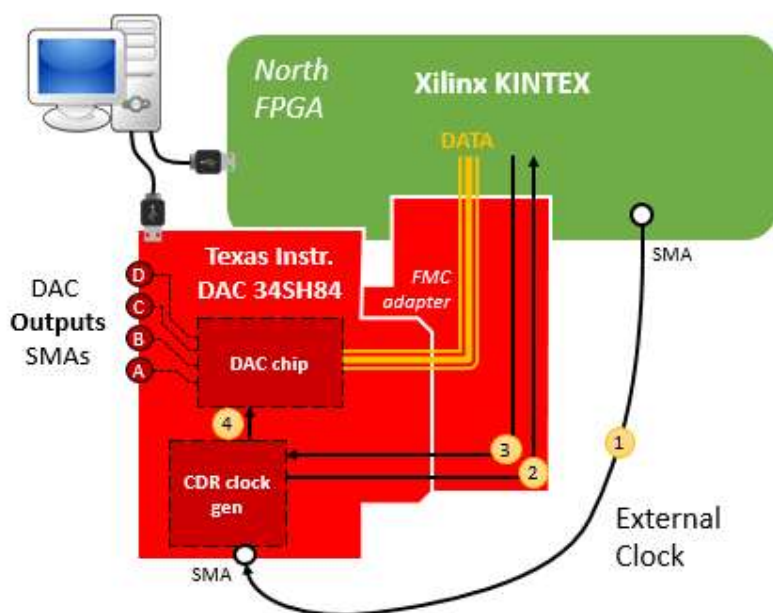


**Εικόνα 66:** α) Φωτογραφία από την ενσωμάτωση του DAC με το FPGA μέσω του προσαρμογέα FMC, β) Παράδειγμα από το γραφικό περιβάλλον σε υπολογιστή για τον χειρισμό του DAC

Παρόλο που η τιμή των ρευμάτων πρέπει να μεταβάλλεται ανά χρονοθυρίδα, δηλαδή κάθε 210  $\mu\text{s}$ , η αρχική ρύθμιση του DAC (AC-coupled) δεν ήταν κατάλληλη για την προτεινόμενη αρχιτεκτονική. Έτσι, έγινε μια μελέτη ώστε να υπάρξει μια ειδικά προσαρμοσμένη διαμόρφωση της συγκεκριμένης συσκευής DAC, σύμφωνα με την οποία επιβάλλεται να μετατραπούν οι διαφορικές έξοδοι μέσω της αφαίρεσης των RF balloons. Το αποτέλεσμα αυτής της προσαρμοσμένης τροποποίησης ήταν ότι οι τέσσερις έξοδοι ρεύματος είναι πλέον DC-coupled και μονής εξόδου, που όμως έχει σαν αποτέλεσμα τη δραστική μείωση του δυναμικού εύρους των τιμών ρεύματος που πλέον διακυμαίνονται από 0,1 mA έως 13,6 mA.

Η ενσωμάτωση του DAC με την πλακέτα FPGA καθώς και όλες οι συνδέσεις που αφορούν στον μεταξύ τους συγχρονισμό (clock – CLK) και τα σήματα ελέγχου που απαιτούνται για την επικοινωνία μεταξύ των δύο μονάδων φαίνονται στην Εικόνα 67. Η μεταβίβαση της ψηφιακής πληροφορίας που περιέχει τις οδηγίες για τις στάθμες ρεύματος σε κάθε έξοδο

γίνεται με αμφίδρομο τρόπο μέσω των κεφαλίδων FMC, αφού διαθέτει 160 ακίδες εισόδου/εξόδου (I/O pins). Το FPGA μπορεί να θεωρηθεί ως κύρια μονάδα συγχρονισμού, καθώς παρέχει το κύριο CLK μέσω ενός εξωτερικού καλωδίου SMA.



**Εικόνα 67:** Οι συνδέσεις μεταξύ του FPGA και του DAC

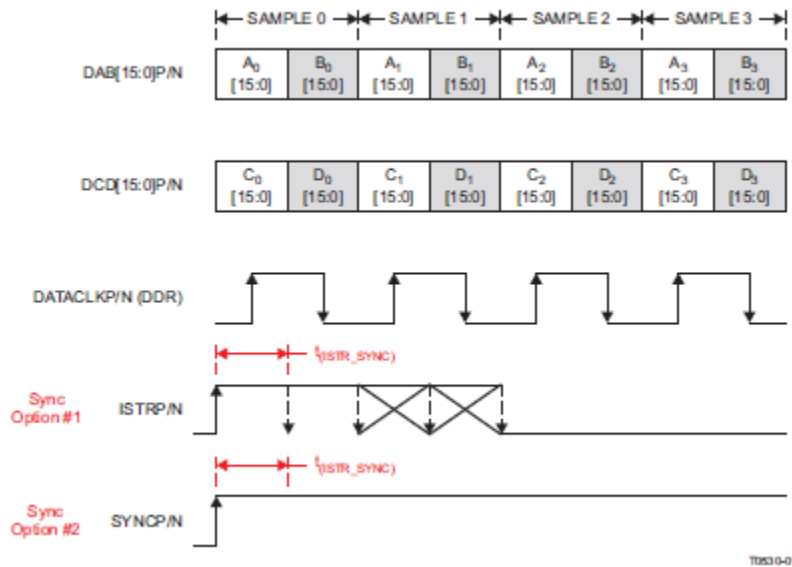
Επίσης, ο DAC είναι εξοπλισμένος με ενσωματωμένο κύκλωμα διανομής ρολογιού (Clock Distribution Circuit – CDC) που επιτρέπει ένα ευρύ φάσμα συχνοτήτων λειτουργίας. Το CDC είναι υπεύθυνο για τη δημιουργία των σημάτων ρολογιού τόσο για τη λειτουργία DAC όσο και για το συγχρονισμό του DAC με το FPGA. Λαμβάνει ως είσοδο ένα ρολόι αναφοράς μέσω ενός καλωδίου SMA - σε αυτήν την περίπτωση ένα σήμα CLK 100 MHz που παράγεται από το FPGA (Εικόνα 67 - σύνδεση 1).

Το CDC δημιουργεί ένα σήμα CLK 50 MHz, κλειδωμένο στο ρολόι εισόδου αναφοράς, το οποίο τροφοδοτείται στο DAC και ορίζει τη συχνότητα λειτουργίας στα 50 MSa/s (Εικόνα 67- σύνδεση 4).

Το ίδιο σήμα CLK μεταβιβάζεται πίσω στο FPGA μέσω του προσαρμογέα FMC για τον έλεγχο του ρυθμού επικοινωνίας καθώς και για να διασφαλιστεί ότι δεν υπάρχει μετατόπιση φάσης μεταξύ του FPGA και του DAC (Εικόνα 67 - σύνδεση 2). Το CLK χρησιμοποιείται από το υποσύστημα FPGA που είναι υπεύθυνο για την αποστολή των σημάτων «DATA» στο DAC. Αυτά τα σήματα είναι στην πραγματικότητα τέσσερις δίαυλοι ελέγχου 16-bit, με κάθε δίαυλο να ορίζει την τιμή του ρεύματος σε κάθε μία από τις τέσσερις εξόδους του DAC.

### 3.2.1 Διεπαφή επικοινωνίας μεταξύ FPGA και DAC

Το πρωτόκολλο επικοινωνίας μεταξύ του FPGA και του DAC φαίνεται στην Εικόνα 68. Το FPGA προκειμένου να επιτύχει σύγχρονη διασύνδεση στέλνει στον DAC τέσσερα σήματα, χρησιμοποιώντας ένα κοινό σήμα ρολογιού (CLK).



**Εικόνα 68:** Η διεπαφή επικοινωνίας μεταξύ του FPGA και του DAC

Η διεπαφή λειτουργεί σε διπλάσιο ρυθμό δεδομένων (double data rate - DDR), που σημαίνει ότι δύο bit από κάθε σήμα μπορούν να μεταδοθούν σε κάθε περίοδο ρολογιού CLK, και πιο συγκεκριμένα ένα σήμα στην άνοδο και ένα στην πτώση του ρολογιού. Πιο συγκεκριμένα, η συνολική διεπαφή επικοινωνίας βασίζεται στα ακόλουθα σήματα:

- “*DATA\_CLK*”, το ρολόι CLK που παράγεται από το CDC και τροφοδοτείται στο FPGA, με φάση μετατοπισμένη κατά 90°. Είναι το σύγχρονο σήμα ρολογιού CLK της διεπαφή (Εικόνα 67 - σύνδεση 3).
- “*AB*”, ένας δίαυλος επικοινωνίας 16 bit (32 ακίδων εφόσον χρησιμοποιείται διαφορετική σηματοδότηση) για σήματα A και B, που μεταφέρει τις τιμές για τις εξόδους A και B κατά τη διάρκεια της ακμής ανόδου και καθόδου του ρολογιού «*DATA\_CLK*» αντίστοιχα.
- “*CD*”, ένας δίαυλος επικοινωνίας 16 bit (32 ακίδων εφόσον χρησιμοποιείται διαφορετική σηματοδότηση) για σήματα C και D, που μεταφέρει τις τιμές για τις εξόδους C και D κατά τη διάρκεια της ακμής ανόδου και καθόδου του ρολογιού «*DATA\_CLK*» αντίστοιχα

- “SYNC”, ένα σήμα 1 bit που χρησιμοποιείται ως σημαία (flag) που εκκινεί την επικοινωνία μεταξύ του master (FPGA) και του slave (DAC).

Για να υποστηρίξει τη λειτουργία DDR, η μονάδα ελέγχου FPGA χρησιμοποιεί το ρολόι των 100 MHz που τροφοδοτείται στο CDC ως εσωτερικό κύριο CLK. Η ταχύτητα που προκύπτει για κάθε έναν από τους δύο διαύλους 16-bit (“AB” και “CD”) είναι 100 MSa/s, ενώ ο κάθε δίαυλος μεταφέρει ψηφιακά δεδομένα στο DAC για δύο κανάλια εναλλάξ, δηλαδή A και B ή C και D αντίστοιχα.

Το FPGA, ως μονάδα ελέγχου, έχει σχεδιαστεί έτσι ώστε σε κάθε κύκλο ρολογιού CLK να μπορεί να ρυθμιστεί η στάθμη των σημάτων A, B, C και D σε διαφορετική τιμή, ανάλογα με το μήκος κύματος που θα πρέπει να συντονιστεί το λέιζερ. Ωστόσο, δεδομένου ότι αυτές οι τιμές ελέγχουν τα αντίστοιχα ρεύματα εξόδου του DAC, πρέπει να ενημερώνονται μόνο κάθε 200 μs, δηλαδή κατά τη διάρκεια του χρόνου παύσης μεταξύ δύο πακέτων. Το εάν και ποιες από τις τιμές θα ενημερωθούν εξαρτάται από τον προγραμματισμό της προτεινόμενης αρχιτεκτονικής, προσδιορίζοντας π.χ. εάν το λέιζερ θα συντονιστεί σε διαφορετικό μήκος κύματος στοχεύοντας διαφορετικό ΤοR προορισμού.

Δεδομένου ότι το λέιζερ χρειάζεται τρία ρεύματα για να συντονιστεί στο επιθυμητό μήκος κύματος, χρησιμοποιούνται οι εξοδοί A, B και C του DAC ενώ η έξοδος D έχει ρυθμιστεί στην ίδια τιμή με την έξοδο C προκειμένου να χρησιμοποιηθεί για σκοπούς παρακολούθησης στον παλμογράφο. Τα ρεύματα που δίνει ο DAC στις θύρες εξόδου A, B και C προσαρμόστηκαν σύμφωνα με τον χάρτη συντονισμού μήκους κύματος του λέιζερ, υποδεικνύοντας τους συνδυασμούς ρεύματος οδήγησης για καθένα από τα 80 μήκη κύματος. Η διαδικασία χαρτογράφησης πραγματοποιήθηκε εκ νέου έτσι ώστε να δημιουργηθεί ένας νέος πίνακας τιμών (Πίνακας 7) για κάθε επιθυμητό μήκος κύματος που να αντιστοιχίζεται με τις ψηφιακές τιμές του DAC.

**Πίνακας 7:** Χαρτογράφηση σημάτων A, B και C που ελέγχουν τα ρεύματα DAC, ανά μήκος κύματος εξόδου του ρυθμιζόμενου λέιζερ.

ITU grid Κανάλι	Συχνότητα (GHz)	Μήκος Κύματος (nm)	Έξοδος A (Φάση)	Έξοδος B (αριστερός αν/ρας)	Έξοδος C (δεξιός αν/ρας)
1	195.900	1530.334	44300	35500	24893
2	195.850	1530.725	52600	33167	23571
3	195.800	1531.116	36000	31933	22714
4	195.750	1531.507	42350	28400	21192
5	195.700	1531.898	43000	26967	55143
6	195.650	1532.290	53250	25036	19458
7	195.600	1532.681	39100	23250	44500

8	195.550	1533.073	42000	22500	41100
9	195.500	1533.465	47000	21000	36300
10	195.450	1533.858	35500	64750	33633
11	195.400	1534.250	38000	58125	31233
12	195.350	1534.643	29500	50950	29000
13	195.300	1535.036	51100	45000	27300
14	195.250	1535.429	39000	43000	25500
15	195.200	1535.822	45800	38500	23500
16	195.150	1536.216	35550	34933	21923
17	195.100	1536.609	52250	33100	61722
18	195.050	1537.003	37150	30800	56792
19	195.000	1537.397	46050	27133	47656
20	194.950	1537.792	36000	25833	44500
21	194.900	1538.186	25000	24143	40533
22	194.850	1538.581	29000	23000	37000
23	194.800	1538.976	48400	21923	34933
24	194.750	1539.371	34000	20500	31200
25	194.700	1539.766	59200	63000	29000
26	194.650	1540.162	25250	55071	26933
27	194.600	1540.557	52600	48000	25400
28	194.550	1540.953	43000	44469	23643
29	194.500	1541.349	29500	41167	21808
30	194.450	1541.746	44300	37406	59900
31	194.400	1542.142	51250	35200	54786
32	194.350	1542.539	36000	32100	48800
33	194.300	1542.936	43950	30000	44500
34	194.250	1543.333	52250	27321	40300
35	194.200	1543.730	23500	25357	36567
36	194.150	1544.128	45600	23500	33033
37	194.100	1544.526	47800	22000	31500
38	194.050	1544.924	36500	21500	28500
39	194.000	1545.322	43850	20000	26100
40	193.950	1545.720	47000	19200	24200
41	193.900	1546.119	23500	53000	23400
42	193.850	1546.518	53750	48000	22000
43	193.800	1546.917	33400	44500	62650

44	193.750	1547.316	37000	39667	53536
45	193.700	1547.715	45300	36800	49033
46	193.650	1548.115	53150	34100	45036
47	193.600	1548.515	35000	30733	40300
48	193.550	1548.915	46000	28167	36467
49	193.500	1549.315	35000	25933	33400
50	193.450	1549.715	29500	24857	31467
51	193.400	1550.116	44500	23000	29000
52	193.350	1550.517	50300	21900	26000
53	193.300	1550.918	37000	20600	24800
54	193.250	1551.319	44750	19500	23000
55	193.200	1551.721	52600	18300	21800
56	193.150	1552.122	37000	53133	63222
57	193.100	1552.524	45350	47625	55393
58	193.050	1552.926	34200	44100	50033
59	193.000	1553.329	35000	39067	44786
60	192.950	1553.731	45800	35167	40433
61	192.900	1554.134	35500	32433	37033
62	192.850	1554.537	39750	29467	33733
63	192.800	1554.940	43400	27600	31800
64	192.750	1555.343	49000	26500	29500
65	192.700	1555.747	38650	24000	27000
66	192.650	1556.151	43800	22200	25000
67	192.600	1556.555	53050	20900	23200
68	192.550	1556.959	46000	20000	21400
69	192.500	1557.363	45550	18700	20300
70	192.450	1557.768	44400	59817	59208
71	192.400	1558.173	37800	52429	51067
72	192.350	1558.578	41650	48167	46433
73	192.300	1558.983	33500	44344	42300
74	192.250	1559.389	36000	39233	38267
75	192.200	1559.794	43150	34967	34400
76	192.150	1560.200	51250	31633	31333
77	192.100	1560.606	38500	28700	28667
78	192.050	1561.013	44750	26714	26679
79	192.000	1561.419	32500	24700	24700
80	191.950	1561.826	55000	23800	23800

### 3.2.2 Χαρακτηρισμός προσαρμοζόμενου πομπού μετά την ενσωμάτωση FPGA και DAC

#### 3.2.2.1 Ανανεωμένη χαρτογράφηση των μηκών κύματος του λέιζερ

Όπως περιγράφεται και στην ενότητα 2.1, ο πομπός της προτεινόμενης αρχιτεκτονικής βασίζεται στον γρήγορο συντονισμό μήκους κύματος ενός προσαρμοζόμενου λέιζερ. Υπενθυμίζεται ότι το λέιζερ που επιλέχθηκε για αυτό το σκοπό είναι ένα λέιζερ της Finisar (S7500 MG-Y), το οποίο ελέγχεται από πέντε ρεύματα εισόδου που εφαρμόζονται σε πέντε αντίστοιχα τμήματα: SOA, κέρδος, φάση, αριστερός ανακλαστήρας και δεξιός ανακλαστήρας. Δύο από αυτά τα ρεύματα αφορούν την ενίσχυση του λέιζερ, τα τμήματα κέρδους (GAIN) και SOA, και μπορούν να διατηρηθούν σταθερά και επομένως χρησιμοποιούνται απλές πηγές σταθερού ρεύματος. Τα υπόλοιπα τρία ρεύματα εφαρμόζονται στα τμήματα φάσης και αριστερού/δεξιού ανακλαστήρα και χρησιμεύουν για τη ρύθμιση του μήκους κύματος εκπομπής του λέιζερ. Για να ενεργοποιηθεί ο γρήγορος συντονισμός μήκους κύματος, τα τρία τελευταία ρεύματα ελέγχονται από μια μονάδα μετατροπεία ψηφιακού σε αναλογικό ρεύμα (DAC34SH84) της Texas Instruments (ενότητα 3.1).

Λόγω του γεγονότος ότι το εύρος των ρευμάτων που αποδίδει ο DAC ήταν περιορισμένο (0,1 mA - 13,6 mA), έπειτα από την τροποποίηση που έγινε στην πλακέτα του, κρίθηκε απαραίτητο να επανεκτιμηθούν οι τιμές του ρεύματος συντονισμού και για τα 80 μήκη κύματος χωρίς να υπερβαίνεται η μέγιστη ικανότητα του DAC. Ο Πίνακας 8 παρουσιάζει τον ενημερωμένο πίνακα τιμών ρευμάτων συντονισμού του λέιζερ.

**Πίνακας 8:** Ενημερωμένος πίνακας τιμών ρευμάτων συντονισμού του λέιζερ, που δείχνει τα αντίστοιχα ρεύματα για τα τμήματα φάσης, αριστερού και δεξιού ανακλαστήρα, αντίστοιχα.

*Η θερμοκρασία κατά τις μετρήσεις ήταν σταθερή στους 25°C και το ρεύμα τόσο για τα τμήματα GAIN όσο και για το SOA ήταν στα 40 mA.*

ITU grid Κανάλι	Συχνότητα (GHz)	Μήκος Κύματος (nm)	Φάση (mA)	Αριστερός Αν/ρας (mA)	Δεξιός Αν/ρας (mA)
1	195.900	1530.334	2.91	2.16	5.29
2	195.850	1530.725	4.57	13.39	4.43
3	195.800	1531.116	5.14	1.54	4.22
4	195.750	1531.507	2.54	1.13	3.17
5	195.700	1531.898	2.67	11.15	2.75

6	195.650	1532.290	4.77	0.69	2.20
7	195.600	1532.681	1.91	8.01	1.69
8	195.550	1533.073	2.65	6.98	1.48
9	195.500	1533.465	4.22	5.53	1.08
10	195.450	1533.858	1.24	4.73	13.27
11	195.400	1534.250	1.86	4.01	11.93
12	195.350	1534.643	3.55	3.35	9.96
13	195.300	1535.036	4.31	2.85	8.17
14	195.250	1535.429	1.89	2.33	7.56
15	195.200	1535.822	3.23	1.76	6.20
16	195.150	1536.216	1.24	1.33	5.12
17	195.100	1536.609	4.57	12.73	4.57
18	195.050	1537.003	1.52	11.60	3.88
19	195.000	1537.397	3.28	8.97	2.80
20	194.950	1537.792	5.14	8.01	2.42
21	194.900	1538.186	2.14	6.81	1.94
22	194.850	1538.581	3.23	5.74	1.62
23	194.800	1538.976	3.76	5.12	1.33
24	194.750	1539.371	2.03	4.01	0.95
25	194.700	1539.766	5.97	3.35	12.97
26	194.650	1540.162	3.21	2.74	11.13
27	194.600	1540.557	4.62	2.28	9.08
28	194.550	1540.953	2.23	1.80	8.00
29	194.500	1541.349	3.09	1.30	7.00
30	194.450	1541.746	2.93	12.35	5.86
31	194.400	1542.142	4.34	11.05	5.20
32	194.350	1542.539	1.66	9.31	4.27
33	194.300	1542.936	2.86	8.01	3.65
34	194.250	1543.333	4.55	6.74	2.86
35	194.200	1543.730	1.86	5.61	2.29
36	194.150	1544.128	3.09	4.55	1.76
37	194.100	1544.526	3.64	4.09	1.35
38	194.050	1544.924	1.54	3.20	1.21
39	194.000	1545.322	2.84	2.50	0.82
40	193.950	1545.720	3.48	1.96	0.63
41	193.900	1546.119	2.84	1.73	10.55



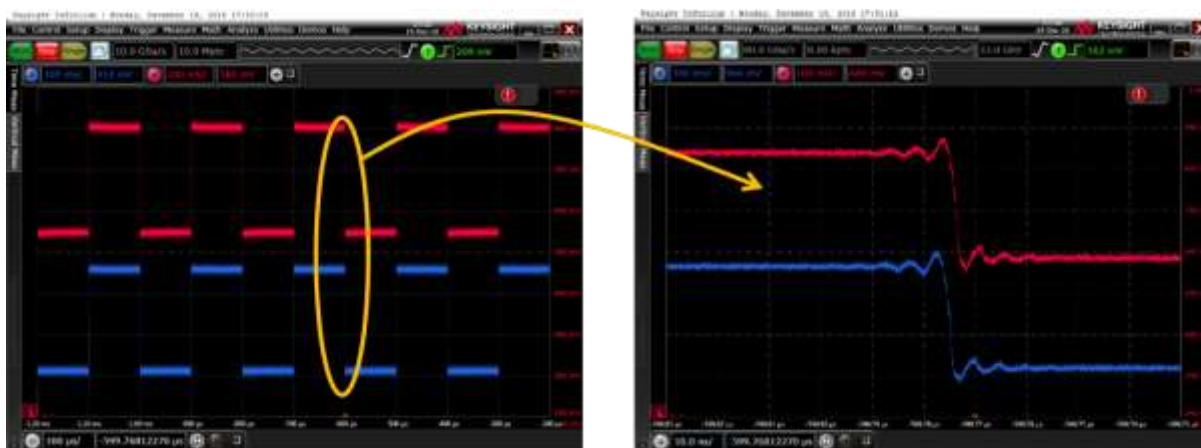
42	193.850	1546.518	4.85	1.35	9.08
43	193.800	1546.917	0.97	12.90	8.01
44	193.750	1547.316	1.94	10.71	6.55
45	193.700	1547.715	3.14	9.38	5.68
46	193.650	1548.115	4.73	8.18	4.87
47	193.600	1548.515	2.05	6.74	3.86
48	193.550	1548.915	3.27	5.58	3.10
49	193.500	1549.315	1.13	4.66	2.45
50	193.450	1549.715	1.93	4.08	2.15
51	193.400	1550.116	2.97	3.35	1.62
52	193.350	1550.517	4.15	2.47	1.32
53	193.300	1550.918	1.78	2.14	0.98
54	193.250	1551.319	3.02	1.62	0.70
55	193.200	1551.721	4.62	1.30	0.41
56	193.150	1552.122	1.63	13.01	10.59
57	193.100	1552.524	3.14	11.22	8.96
58	193.050	1552.926	0.99	9.68	7.89
59	193.000	1553.329	2.05	8.11	6.37
60	192.950	1553.731	3.23	6.78	5.19
61	192.900	1554.134	1.20	5.75	4.37
62	192.850	1554.537	2.03	4.76	3.49
63	192.800	1554.940	2.75	4.18	2.94
64	192.750	1555.343	3.88	3.50	2.62
65	192.700	1555.747	1.82	2.76	1.90
66	192.650	1556.151	2.83	2.19	1.40
67	192.600	1556.555	4.71	1.68	1.05
68	192.550	1556.959	1.94	1.18	0.82
69	192.500	1557.363	3.18	0.90	0.51
70	192.450	1557.768	2.95	11.70	12.23
71	192.400	1558.173	1.66	9.99	10.39
72	192.350	1558.578	2.40	8.66	9.21
73	192.300	1558.983	0.89	7.35	7.96
74	192.250	1559.389	1.40	6.13	6.42
75	192.200	1559.794	2.70	4.96	5.13
76	192.150	1560.200	4.34	4.04	4.13
77	192.100	1560.606	6.18	3.25	3.26

78	192.050	1561.013	3.02	2.67	2.68
79	192.000	1561.419	4.43	2.10	2.10
80	191.950	1561.826	5.52	1.84	1.84

### 3.2.2.2 Χρόνος απόκρισης και ρύθμισης του λέιζερ στο νέο υποσύστημα FPGA/DAC

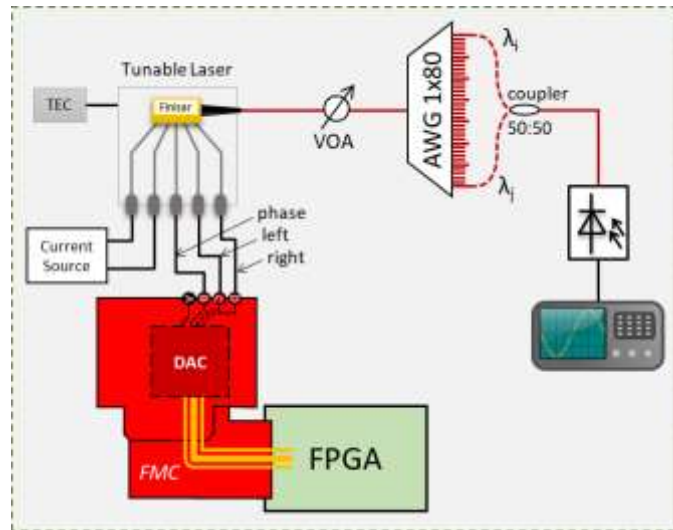
Στην ενότητα 2.1.4 μελετήθηκε ο χρόνος μεταγωγής του προσαρμοζόμενου λέιζερ, ενώ μετρήθηκε να είναι περίπου  $\sim 8$  ns, ξεπερνώντας τις προδιαγραφές στόχου των 50 ns που επιβάλλεται από την προτεινόμενη αρχιτεκτονική δικτύου. Είναι σημαντικό να αναφερθεί ότι για τον υπολογισμό του συνολικού χρόνου μεταγωγής του πομπού στο ToR switch, πρέπει να ληφθεί υπόψη η συμπεριφορά του DAC, δηλαδή ο χρόνος ο οποίος απαιτείται για τη μετάβαση από μια στάθμη ρεύματος σε μία άλλη.

Προς τούτο, το FPGA προγραμματίζεται έτσι ώστε να ρυθμίσει τον DAC να μεταβαίνει από μία στάθμη σε μία άλλη περιοδικά. Οι κυματομορφές που φαίνονται με χρώμα κόκκινο και μπλε στην Εικόνα 69, αντιπροσωπεύουν δύο ηλεκτρικές εξόδους του DAC. Εστιάζοντας σε μια μεταβολή δύο σταθμών είναι δυνατό να μετρηθεί ο χρόνος μετάβασης. Η μετάβαση μετρήθηκε να είναι σχεδόν 10 ns και συμφωνεί με τις προδιαγραφές του κατασκευαστή του DAC.



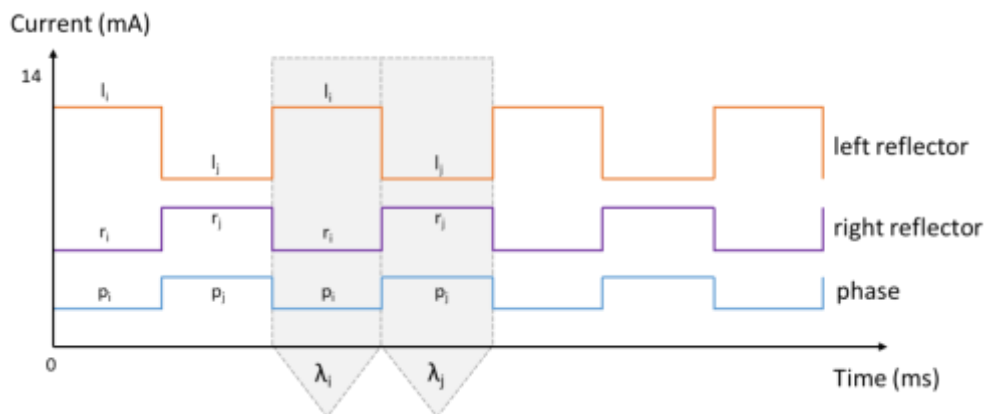
*Εικόνα 69: Ηλεκτρικά σήματα που παράγονται από το DAC. Παρατηρείται ότι η μετάβαση από μια στάθμη σε μια άλλη διαρκεί σχεδόν 10 ns.*

Μετά την αξιολόγηση της συμπεριφοράς του DAC, εφαρμόστηκε η πειραματική διάταξη που φαίνεται στην Εικόνα 70, προκειμένου να διερευνηθεί ο συνολικός χρόνος μεταγωγής και συντονισμού του πομπού από ένα οποιοδήποτε κανάλι σε ένα άλλο.



**Εικόνα 70:** Η Πειραματική διάταξη που χρησιμοποιείται για τον προσδιορισμό της συνολικής ταχύτητας συντονισμού του προσαρμοζόμενου πομπού.

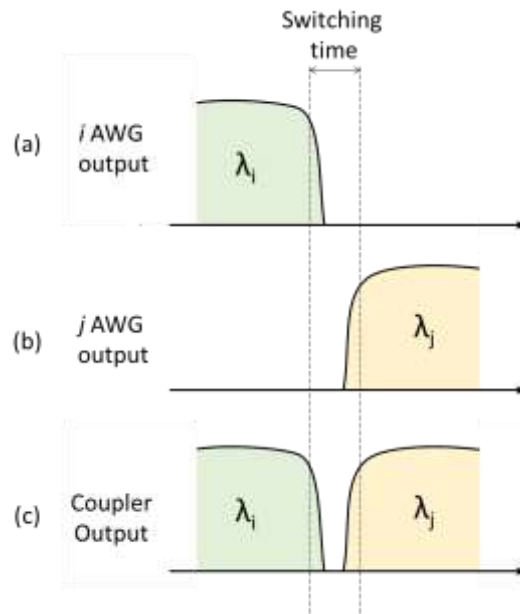
Η μεθοδολογία που ακολουθείτε για τη μέτρηση του χρόνου μεταγωγής μεταξύ δύο μηκών κύματος είναι η ακόλουθη: Το σύστημα FPGA/DAC «φτιάχνει» τα ρεύματα που ρυθμίζουν το λέιζερ έτσι ώστε να εκπέμπει εναλλάξ συνεχές φως σε δύο μήκη κύματος ( $\lambda_i$  και  $\lambda_j$ ), όπως φαίνεται στα σήματα στην Εικόνα 71. Έτσι, όταν ο DAC δίνει τα ρεύματα  $l_i$  (left reflector),  $r_i$  (right reflector) and  $p_i$  (phase section) το λέιζερ εκπέμπει στο  $\lambda_i$ . Ομοίως, όταν το DAC παρέχει τα ρεύματα  $l_j$ ,  $r_j$  και  $p_j$  το λέιζερ συντονίζεται στο  $\lambda_j$ .



**Εικόνα 71:** Τα ηλεκτρικά σήματα που παράγονται από το DAC για τη ρύθμιση μήκους κύματος του λέιζερ μεταξύ  $\lambda_i$  και  $\lambda_j$ . Τα πορτοκαλί, μοβ και μπλε ίχνη αντιπροσωπεύουν τα

ρεύματα που εγχέονται στον αριστερό ανακλαστήρα του λέιζερ ( $l_i$  και  $l_j$ ), στον δεξιό ανακλαστήρα ( $r_i$  και  $r_j$ ) και στο τμήμα φάσης ( $p_i$  και  $p_j$ ) αντίστοιχα.

Κάθε ζεύγος μηκών κύματος που εκπέμπεται από το λέιζερ αποπολυπλέκεται μέσω ενός AWG  $1 \times 80$  προκειμένου να προσδιοριστεί πότε η έξοδος του λέιζερ έχει ρυθμιστεί στο  $\lambda_i$  και πότε στο  $\lambda_j$ , έτσι ώστε να αποκαλυφθεί η πραγματική διάρκεια της μετάβασης (switching time). Η Εικόνα 72 (α) και (β) απεικονίζει τις εξόδους του  $1 \times 80$  AWG για  $\lambda_i$  και  $\lambda_j$  αντίστοιχα. Για λόγους ευκολίας στην εκτέλεση της μέτρησης, χρησιμοποιείται ένας ζεύκτης 50:50 για το συνδυασμό των δύο σημάτων. Ο συνδυασμός αυτών των σημάτων ανιχνεύεται από μια φωτοδίοδο και φαίνεται στην Εικόνα 72 (γ).



**Εικόνα 72:** (α) Έξοδος του  $1 \times 80$  AWG που αντιστοιχεί στο  $\lambda_i$  και (β) έξοδος του  $1 \times 80$  AWG που αντιστοιχεί σε  $\lambda_j$  και (γ) Το συνδυασμένο σήμα όπως ανιχνεύεται από τη φωτοδίοδο και χρησιμοποιείται για τη μέτρηση του χρόνου μεταγωγής για κάθε ζεύγος μηκών κύματος.

Το κανάλι 51 (1550,116 nm) σύμφωνα με το πρότυπο ITU επιλέχθηκε τυχαία ως το κανάλι αναφοράς  $\lambda_i$ . Πραγματοποιήθηκαν ογδόντα συνδυασμοί αλλάζοντας το μήκος κύματος  $\lambda_j$  για  $j=1,2,\dots,80$ . Ο Πίνακας 9 παρουσιάζει τον πειραματικά μετρημένο χρόνο εναλλαγής σε νανοδευτερόλεπτα από το κανάλι 51 σε καθένα από τα υπόλοιπα 79 κανάλια της C μπάντας.

**Πίνακας 9:** Χρόνος εναλλαγής σε νανοδευτερόλεπτα (ns) από το κανάλι 51 (1550,116 nm) σε καθένα από τα υπόλοιπα 79 κανάλια της C μπάντας, σύμφωνα με τα κεντρικά μήκη κύματος που καθορίζονται από το πρότυπο ITU-T.

Κανάλι (ITU grid)	Μήκος Κύματος (nm)	Χρόνος εναλλαγής (ns)	Κανάλι (ITU grid)	Μήκος Κύματος (nm)	Χρόνος εναλλαγής (ns)
1	1530.334	17.2	41	1546.119	12.8
2	1530.725	14.4	42	1546.518	17.9
3	1531.116	19.6	43	1546.917	16.8
4	1531.507	21.5	44	1547.316	14.0
5	1531.898	12.4	45	1547.715	17.6
6	1532.29	18.3	46	1548.115	10.2
7	1532.681	10.0	47	1548.515	13.6
8	1533.073	10.6	48	1548.915	12.2
9	1533.465	11.6	49	1549.315	12.0
10	1533.858	14.6	50	1549.715	8.2
11	1534.25	18.4	51	1550.116	-
12	1534.643	12.4	52	1550.517	14.2
13	1535.036	20.1	53	1550.918	16.8
14	1535.429	18.4	54	1551.319	19.7
15	1535.822	16.6	55	1551.721	17.4
16	1536.216	16.8	56	1552.122	14.4
17	1536.609	11.6	57	1552.524	14.1
18	1537.003	13.2	58	1552.926	14.8
19	1537.397	13.1	59	1553.329	12.4
20	1537.792	10.6	60	1553.731	16.4
21	1538.186	17.2	61	1554.134	18.2
22	1538.581	9.4	62	1554.537	13.6
23	1538.976	10.0	63	1554.940	10.8
24	1539.371	15.6	64	1555.343	12.1
25	1539.766	17.1	65	1555.747	8.9
26	1540.162	19.5	66	1556.151	8.6
27	1540.557	20.1	67	1556.555	12.3
28	1540.953	18.8	68	1556.959	11.2
29	1541.349	22.3	69	1557.363	17.2
30	1541.746	12.4	70	1557.768	15.6
31	1542.142	12.4	71	1558.173	13.8
32	1542.539	11.2	72	1558.578	14.8
33	1542.936	12.2	73	1558.983	12.5
34	1543.333	9.6	74	1559.389	11.8

<b>35</b>	1543.730	16.8
<b>36</b>	1544.128	11.2
<b>37</b>	1544.526	8.1
<b>38</b>	1544.924	8.0
<b>39</b>	1545.322	12.4
<b>40</b>	1545.720	20.3

<b>75</b>	1559.794	15.2
<b>76</b>	1560.200	15.2
<b>77</b>	1560.606	17.2
<b>78</b>	1561.013	16.4
<b>79</b>	1561.419	17.1
<b>80</b>	1561.826	15.3

Οι διακυμάνσεις στο χρόνο εναλλαγής (switching time) που παρατηρούνται στον πίνακα 9 συμφωνούν με τη βιβλιογραφία. Πιο συγκεκριμένα, η αναφορά [46] σχετική με τα δυναμικά χαρακτηριστικά συντονισμού του συγκεκριμένου λέιζερ τύπου MG-Y της FINISAR, για διαφορετικούς τύπους μεταβάσεων μήκους κύματος:

Πιο συγκεκριμένα, η αναφορά [46] αναφέρει τα δυναμικά χαρακτηριστικά συντονισμού του λέιζερ MG-Y, για διαφορετικούς ρυθμούς μεταβάσεων μήκους κύματος:

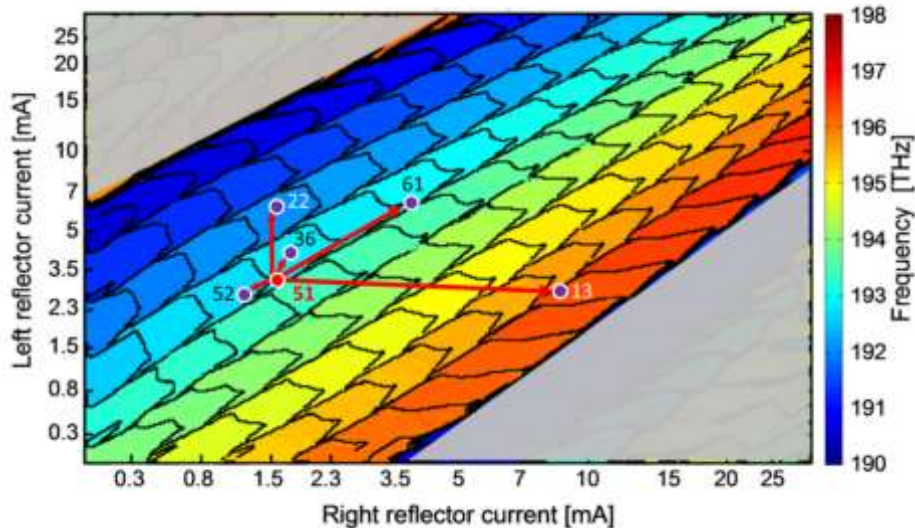
1. εντός του ίδιου ρυθμού,
2. εντός του ίδιου υπερ-ρυθμού σε διαδοχικούς ρυθμούς,
3. εντός του ίδιου υπερ-ρυθμού σε μη διαδοχικούς ρυθμούς,
4. σε διαδοχικούς υπερ-ρυθμούς,
5. σε μη διαδοχικούς υπερ-ρυθμούς

Ο Πίνακας 10 παρουσιάζει τον χρόνο εναλλαγής για ενδεικτικά ζεύγη μηκών κύματος που αντιστοιχούν σε καθέναν από τους παραπάνω τύπους μετάβασης. Επίσης, στο σχήμα 73 απεικονίζονται αυτά τα ζεύγη μηκών κύματος, με το μήκος κύματος αναφοράς να επισημαίνεται με κόκκινο χρώμα. Τα αποτελέσματα συμφωνούν με τα αποτελέσματα που αναφέρονται στην αναφορά [46] και εμφανίζονται στην τέταρτη στήλη του πίνακα 10.

**Πίνακας 10:** Χρόνος συντονισμού μήκους κύματος λέιζερ για κάθε τύπο μετάβασης.

Τύπος μετάβασης	Κανάλια	Χρόνος μετάβασης μαζί με τον χρόνο μετάβασης του DAC	Αναφορά [46] (χωρίς τον χρόνο μετάβασης του DAC)
Μέσα στον ίδιο ρυθμό	51 → 36	11.2 ns	3 ns
Εντός του ίδιου υπερ-ρυθμού σε διαδοχικούς ρυθμούς	51 → 52	14.2 ns	4 ns
Εντός του ίδιου υπερ-ρυθμού σε μη διαδοχικούς ρυθμούς	51 → 61	18.2 ns	7 ns

Σε διαδοχικούς υπερ-ρυθμούς	51 → 22	9.4 ns	0.4 ns
Σε μη διαδοχικούς υπερ-ρυθμούς	51 → 13	20.1 ns	12 ns



*Εικόνα 73: Η συχνότητα εκπομπής ενός λέιζερ MG-Y απεικονίζεται ως συνάρτηση των ρευμάτων του αριστερού και του δεξιού ανακλαστήρα [46]*

### 3.3 Ενσωμάτωση συστημάτων ToR και pod switch

Σε αυτήν την ενότητα παρουσιάζονται τα πειραματικά αποτελέσματα που επιδεικνύουν την επιτυχημένη ενσωμάτωση και διασύνδεση όλων των οπτικών υποσυστημάτων του ToR και του pod switch. Τα αποτελέσματα αυτά αποτέλεσαν σημείο καμπής για την αξιολόγηση των τεχνικών δρομολόγησης και της συνολικής ολοκλήρωσης του συστήματος με τελικό στόχο το πρωτότυπο μίνι data-center της προτεινόμενης αρχιτεκτονικής που θα παρουσιαστεί στο ΚΕΦΑΛΑΙΟ 5.

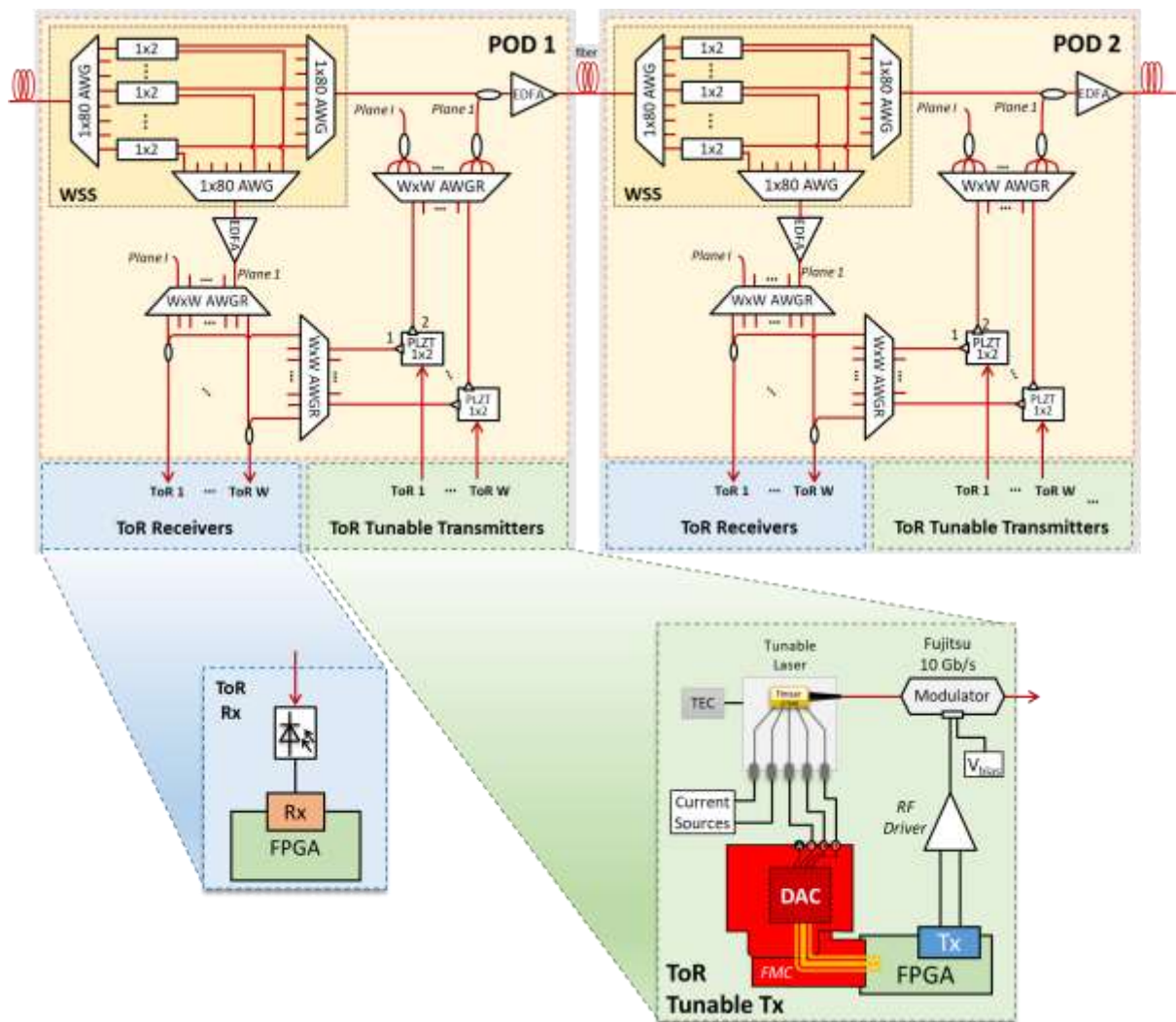
#### 3.3.1 Αξιολόγηση της ενσωμάτωσης των υποσυστημάτων μέσω διαφορετικών τεχνικών δρομολόγησης

Στο ΚΕΦΑΛΑΙΟ 2 παρουσιάστηκαν και αξιολογήθηκαν τα οπτικά υποσυστήματα της υπό διερεύνηση αρχιτεκτονικής δικτύου. Η αξιολόγησή τους εκεί επικεντρώθηκε στο φυσικό στρώμα του δικτύου τόσο σε intra-pod όσο και σε inter-pod σενάριο δρομολόγησης. Ως

επόμενο βήμα, σε αυτήν την ενότητα, παρουσιάζονται παρόμοια σενάρια που αναφέρονται στην επικοινωνία εντός και μεταξύ των pods με την ενσωμάτωση των οπτικών υποσυστημάτων του ToR και το pod switch, συμπεριλαμβανομένων των FPGA που χρησιμοποιούνται για τον έλεγχο των οπτικών στοιχείων και τη διασύνδεση με το επίπεδο ελέγχου (control-plane).

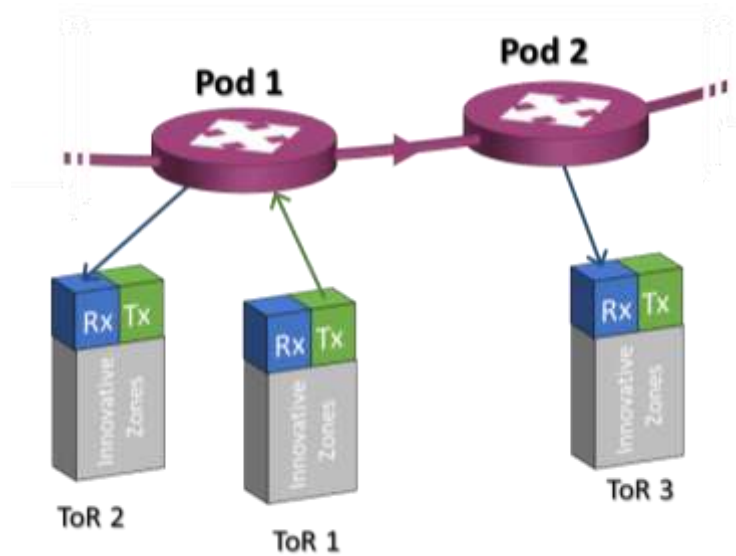
Η Εικόνα 74 απεικονίζει σχηματικά τη διάταξη που περιγράφει το φυσικό επίπεδο της προτεινόμενης οπτικής αρχιτεκτονικής δικτύου. Πιο συγκεκριμένα, απεικονίζεται η σύνδεση δύο διαδοχικών pod μέσω του WDM δακτυλίου. Κάθε pod διαθέτει έναν pod switch που εξυπηρετεί W ToR σε μια τοπολογία αστεριού. Οι ToR είναι εξοπλισμένοι με ρυθμιζόμενα λέιζερ και πομπές και δέκτες (burst-mode) που ελέγχονται από FPGA. Για κάθε ToR πομπό υπάρχει ένας 1×2 οπτικός διακόπτης καθορισμού που βρίσκεται στο pod switch, όπως φαίνεται στην Εικόνα 74 και καθορίζει την δρομολόγηση της ροής εντός ή μεταξύ των pod. Επίσης, ο pod switch διαθέτει παθητικούς οπτικούς δρομολογητές (AWGR) που είναι υπεύθυνοι για τη δρομολόγηση κάθε μήκους κύματος είτε στο ToR προορισμού μέσα στο pod είτε στον οπτικό δακτύλιο.





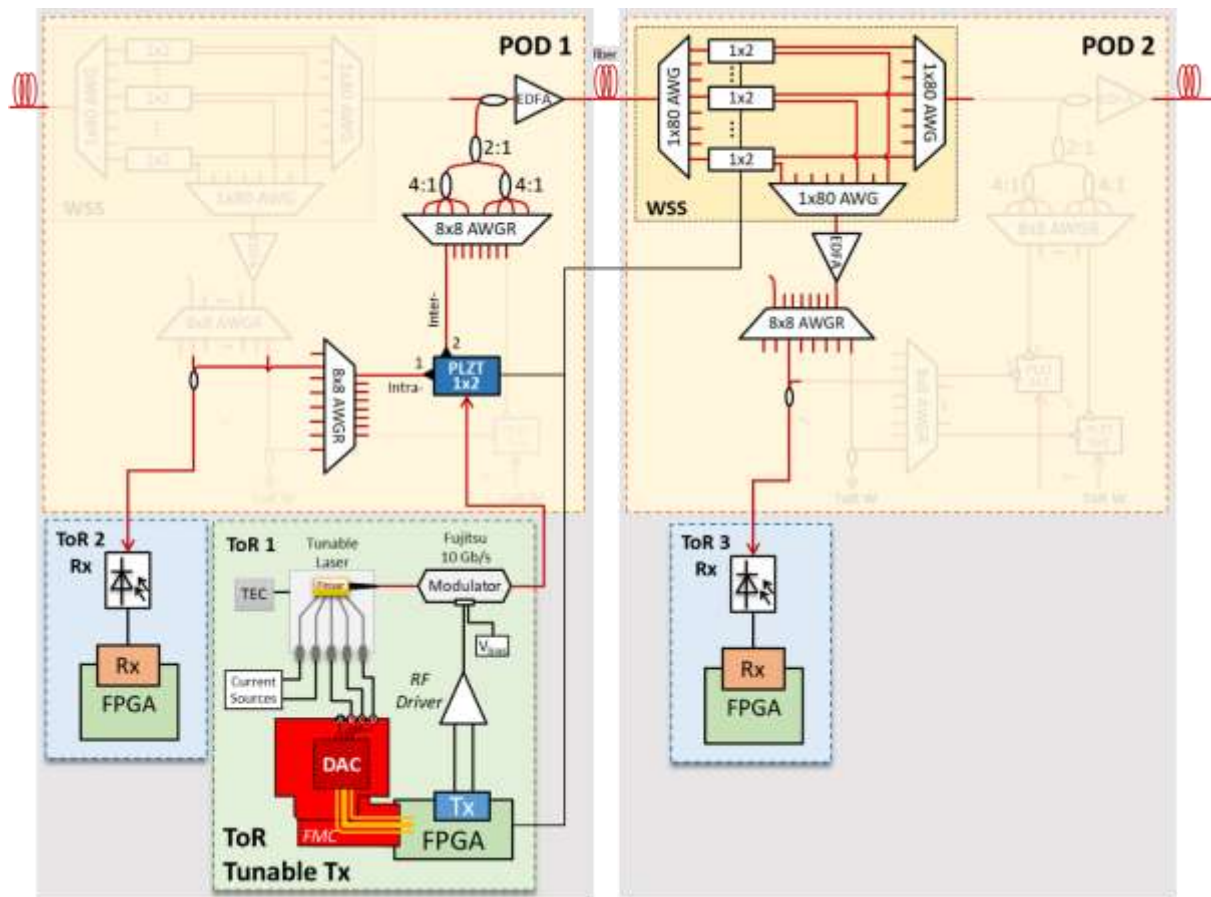
*Εικόνα 74: Σχηματική διάταξη δύο διαδοχικών pod της προτεινόμενης αρχιτεκτονικής*

Ωστόσο, για τις ανάγκες των πειραμάτων αξιολόγησης των σεναρίων εντός και μεταξύ των pod, η διάταξη που εφαρμόστηκε ήταν απλοποιημένη σε σχέση με την διάταξη της εικόνας 74. Όπως φαίνεται και στην Εικόνα 75, η πειραματική διάταξη που περιγράφεται σε αυτήν την ενότητα περιλαμβάνει δύο pod switches (pod 1 και pod 2) και τρεις ToR switches (ToR 1, ToR 2 και ToR 3), έτσι ώστε να μπορούν να πραγματοποιηθούν και να δοκιμαστούν τόσο τα σεναρία επικοινωνίας εντός όσο και μεταξύ των pod, καθώς και οι συνδυασμοί τους.



*Εικόνα 75: Επισκόπηση του απλοποιημένου σχηματικού για τις πειραματικές δοκιμές*

Από την άλλη, η Εικόνα 76 παρουσιάζει την απλοποιημένη πειραματική διάταξη που υλοποιήθηκε. Τα τμήματα που αγνοούνται δεν υλοποιήθηκαν και παρουσιάζονται για λόγους πληρότητας. Πιο συγκεκριμένα, το pod 1 περιέχει όλο το οπτικό υποσύστημα για το χειρισμό της κυκλοφορίας εντός του pod καθώς και της εξερχόμενης κίνησης από το pod 1 προς το pod 2 (τα εισερχόμενα στο pod 1 πακέτα διαχειρίζονται από το σκιασμένο τμήμα που δεν υλοποιήθηκε). Ως εκ τούτου, το ToR 1 χρησιμεύει ως πομπός για την επικοινωνία εντός του pod 1 ή μεταξύ των δύο pods, ενώ το ToR 2 χρησιμεύει μόνο ως δέκτης για την κίνηση εντός του pod. Το pod 2, καθώς και ο υποκείμενος ToR 3 switch, είναι εξοπλισμένοι μόνο με τα υποσυστήματα που είναι απαραίτητα για τη διαχείριση της εισερχόμενης κίνησης εντός του pod.



*Εικόνα 76: Απλοποιημένη πειραματική διάταξη για τη δοκιμή της επικοινωνίας εντός και μεταξύ pod μέσω δυναμικού ελέγχου από FPGA των οπτικών υποσυστημάτων.*

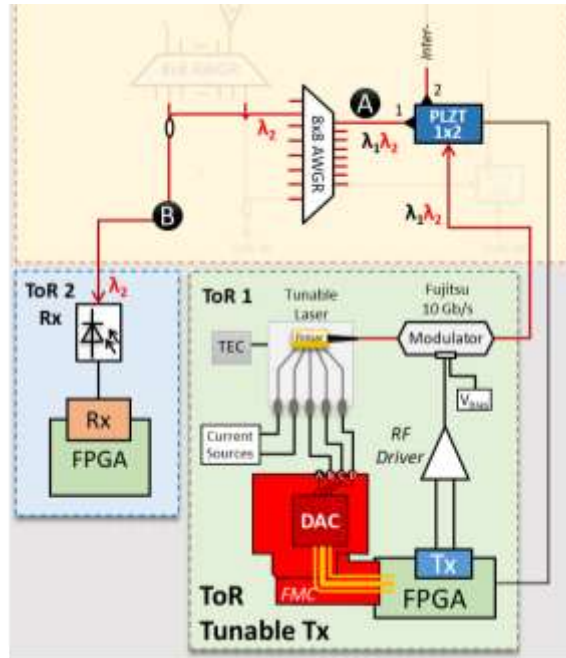
### 3.3.2 Αξιολόγηση μέσω Intra-pod δρομολόγησης

Η πειραματική διάταξη του σχήματος 77 χρησιμοποιήθηκε για την αξιολόγηση της επικοινωνίας μεταξύ κεντρικών υπολογιστών που βρίσκονται εντός του ίδιου pod και διαφορετικών ToRs, δηλαδή του intra-pod σεναρίου επικοινωνίας. Ο προσαρμοζόμενος πομπός (Tunable Tx) εκπέμπει εναλλάξ και συνεχόμενα πακέτα εγγεγραμμένα στα μήκη κύματος  $\lambda_1=1546,91$  nm και  $\lambda_2=1551,72$  nm.

Ο 1x2 οπτικός διακόπτης PLZT, που φαίνεται με σκούρο μπλε χρώμα στο σχήμα 77, είναι ο διακόπτης καθορισμού και διαχωρίζει τα πακέτα που προορίζονται για εντός του pod 1 από τα πακέτα που εξέρχονται του pod 1, όπως φαίνεται στην Εικόνα 78 (Α).

Πιο συγκεκριμένα, η αλληλουχία των πακέτων έχει ως εξής: 8 πακέτα δρομολογούνται εκτός του pod 1, δηλαδή έχουν προορισμό σε κάποιο άλλο pod ενώ 8 πακέτα παραμένουν εντός του pod 1, μέσω των εξόδων 1 και 2 του διακόπτη καθορισμού. Τα πακέτα της intra-pod διαδρομής (έξοδος 1 του οπτικού διακόπτη 1x2) φτάνουν στον ToR προορισμού τους μέσω του 8x8 AWGR. Η Εικόνα 78 (Β) απεικονίζει τα πακέτα που έχουν προορισμό το

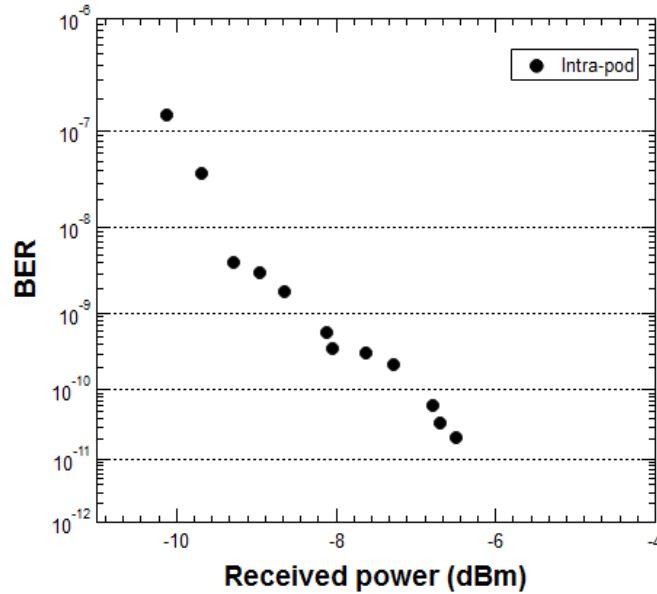
ToR 2 και είναι γραμμένα στο  $\lambda_2$ . Τέλος η Εικόνα 79 παρουσιάζει την αντίστοιχη καμπύλη BER ως συνάρτηση της λαμβανόμενης οπτικής ισχύος. Για λαμβανόμενη ισχύ μεγαλύτερη από  $-6,5$  dBm δεν παρατηρήθηκαν σφάλματα, που αντιστοιχούν σε BER καλύτερο από  $3,10^{-13}$ .



*Εικόνα 77: Πειραματική διάταξη για την αξιολόγηση της ενσωμάτωσης ToR και rod switch σε ένα intra-rod σενάριο δρομολόγησης*



*Εικόνα 78: Στιγμιότυπα των πακέτων που είναι εγγεγραμμένα στα  $\lambda_1$  και  $\lambda_2$  και περνούν από τη διαδρομή intra-rod. Τα σημεία A και B υποδεικνύονται στην πειραματική διάταξη του σχήματος 77.*

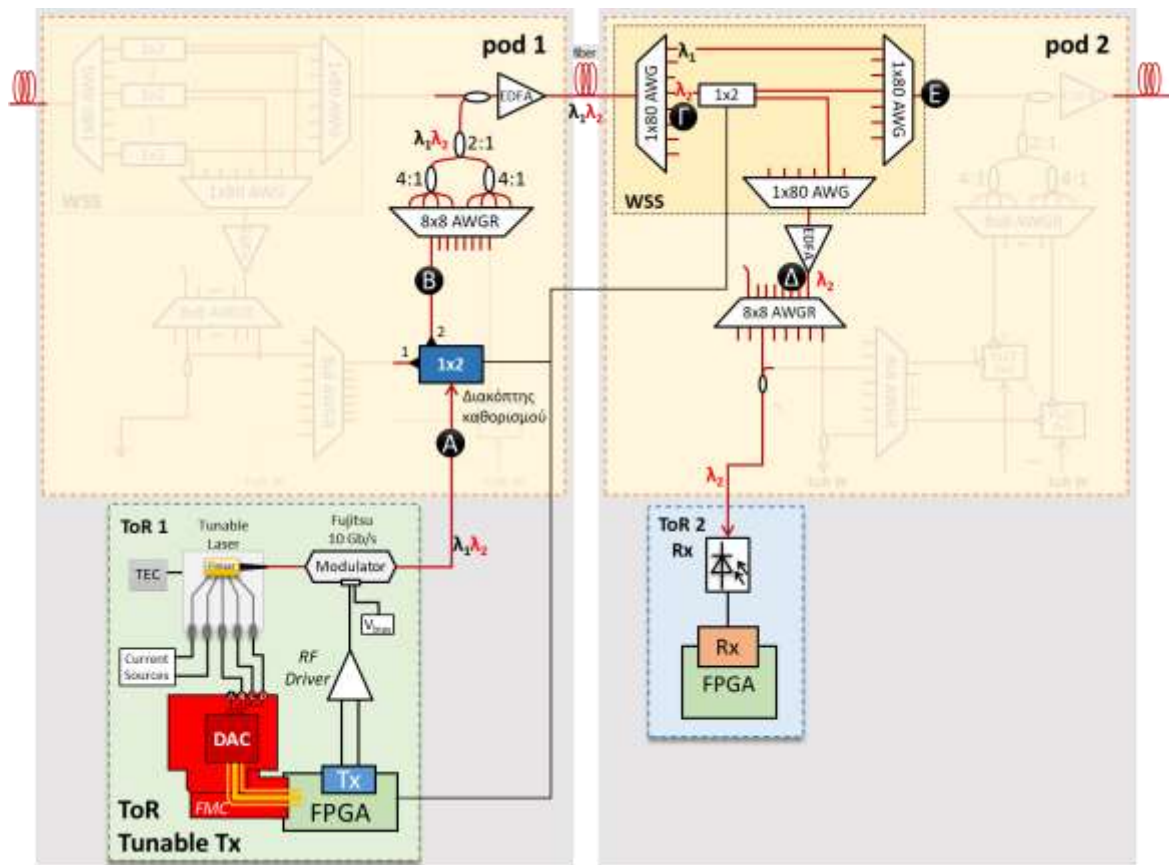


*Εικόνα 79: Το BER ως συνάρτηση της λαμβανόμενης ισχύος για το intra-pod σενάριο*

### 3.3.3 Αξιολόγηση μέσω Inter-pod δρομολόγησης

Αυτή η σειρά πειραμάτων αφορά στην αξιολόγηση της ενσωμάτωσης των υποσυστημάτων ToR και pod switch και περιλαμβάνουν όλα τα επιμέρους υποσυστήματα που περιγράφονται στις προηγούμενες ενότητες του κεφαλαίου 3. Η Εικόνα 80 παρουσιάζει την πειραματική διάταξη που χρησιμοποιήθηκε προκειμένου να επιτευχθεί επικοινωνία μεταξύ δύο pod (inter-pod). Πιο συγκεκριμένα, το ToR 1 που βρίσκεται στο pod 1 δημιουργεί πακέτα εγγεγραμμένα στα μήκη κύματος  $\lambda_1=1546,91$  nm και  $\lambda_2=1551,72$  nm και ο προορισμός τους βρίσκεται στο pod 2. Έτσι, μέσω του  $1 \times 2$  διακόπτη καθορισμού τα πακέτα δρομολογούνται στην έξοδο 2 και σαν αποτέλεσμα εξέρχονται από το pod1 με προορισμό τον οπτικό δακτύλιο, μέσω του οποίου θα φτάσουν στο pod προορισμού. Η οπτική κίνηση των πακέτων ενισχύεται από ένα EDFA που βρίσκεται στην έξοδο του pod 1, πριν εισέλθει στο pod 2 μέσω του διακόπτη επιλογής μήκους κύματος (WSS).

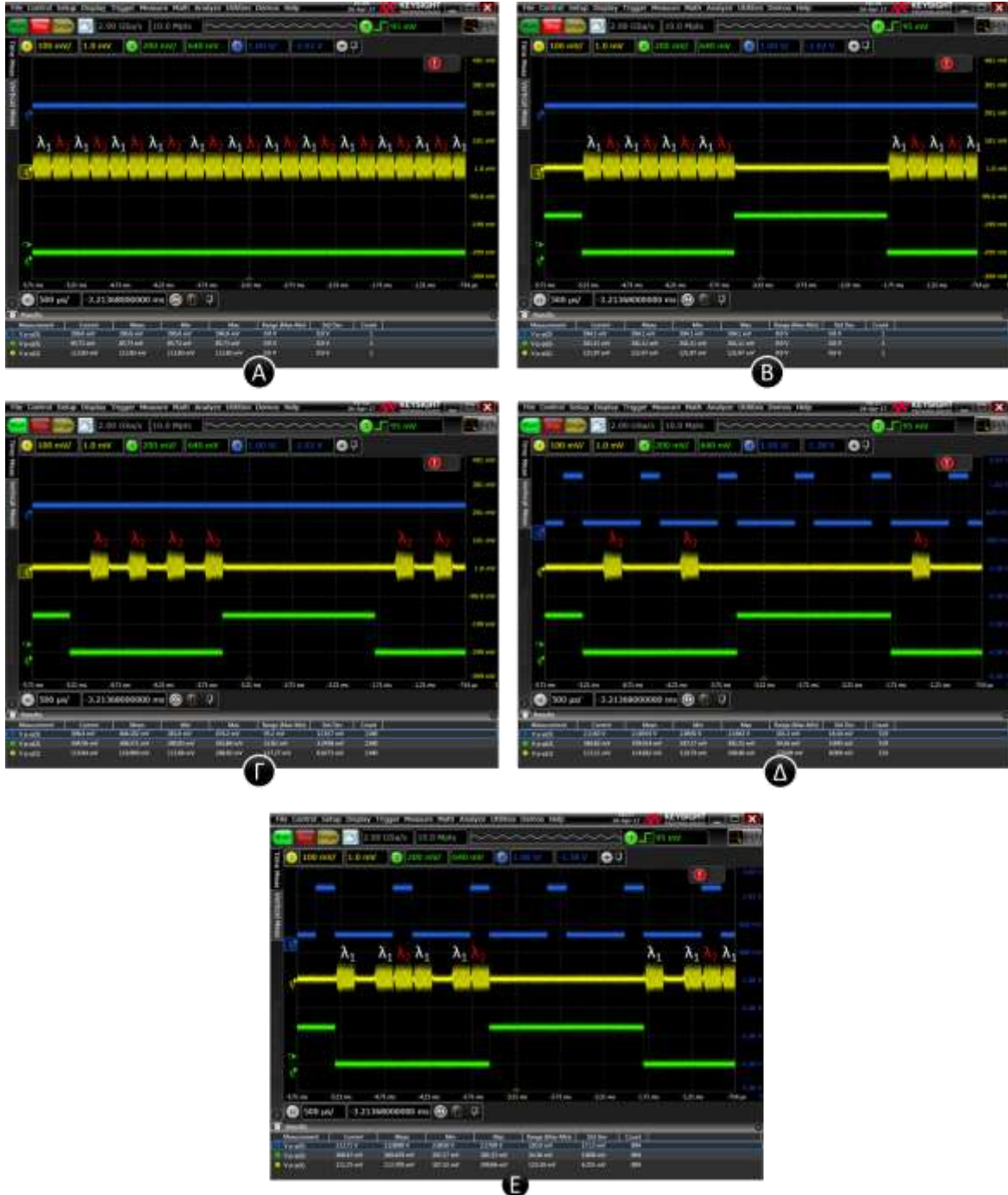
Λόγω των περιορισμένων διαθέσιμων οπτικών διακοπών κατά τη διάρκεια διεξαγωγής του πειράματος, μόνο ένας οπτικός διακόπτης  $1 \times 2$  χρησιμοποιήθηκε για την υλοποίηση του διακόπτη επιλογής μήκους κύματος (WSS) που βρίσκεται στην υποδοχή του pod 2.



**Εικόνα 80:** Πειραματική διάταξη για την αξιολόγηση της ενσωμάτωσης ToR και pod switch σε ένα inter-pod σενάριο δρομολόγησης

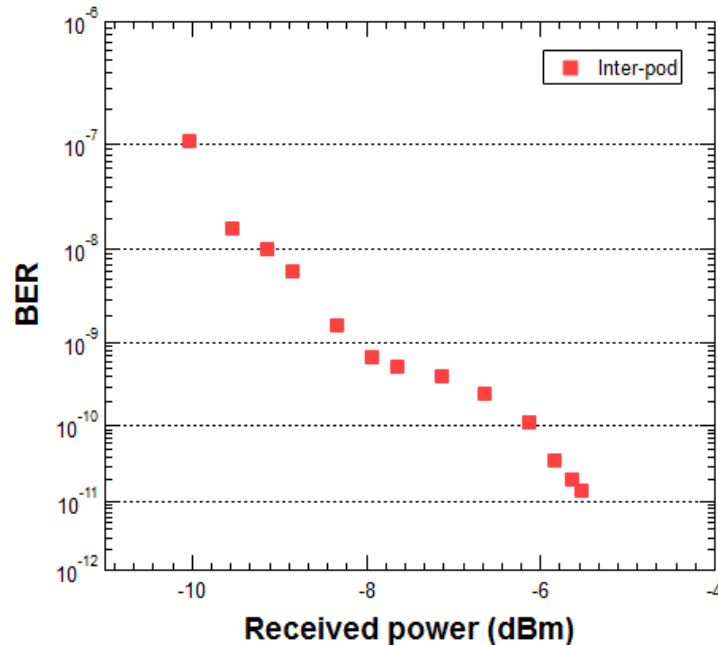
Τα στιγμιότυπα (Α-Δ) που παρουσιάζονται στην Εικόνα 81 δείχνουν τα πακέτα που αντιστοιχούν στα σημεία Α, Β, Γ, Δ και Ε του σχήματος της πειραματικής διάταξης στην Εικόνα 80. Το στιγμιότυπο Α απεικονίζει όλα τα πακέτα που παράγονται από τον προσαρμοσμένο πομποδέκτη του ToR 1 και είναι «γραμμένα» εναλλάξ στα μήκη κύματος  $\lambda_1$  και  $\lambda_2$ . Εν συνεχεία, το στιγμιότυπο Β απεικονίζει εκείνα τα πακέτα που δρομολογούνται στην έξοδο 2 του διακόπτη καθορισμού και αφού περάσουν μέσα από παθητικά οπτικά στοιχεία εισέρχονται στην ίνα του οπτικού δακτυλίου με προορισμό το pod 2. Έπειτα, όλα τα πακέτα εισέρχονται στον διακόπτη επιλογής μήκους κύματος του pod 2, ο οποίος αποφασίζει με βάση το μήκος κύματος ποια πακέτα θα προωθηθούν στον δακτύλιο και ποια θα εισέλθουν στο pod2. Έτσι, το στιγμιότυπο Γ στην Εικόνα 81, απεικονίζει τα πακέτα που είναι γραμμένα στο  $\lambda_2$  και βρίσκονται μέσα στον διακόπτη επιλογής μήκους κύματος (WSS) αλλά πριν από τον επιμέρους 1x2 διακόπτη που θα αποφασίσει για τη δρομολόγησή τους. Κατόπιν και εφόσον εφαρμόζεται το σήμα ελέγχου που φαίνεται με μπλε χρώμα στο στιγμιότυπο Δ, ο 1x2 διακόπτης «ρίχνει» τα δύο από τα τέσσερα πακέτα του  $\lambda_2$  μέσα στο pod 2, ενώ προωθεί τα εναπομείναντα πακέτα του  $\lambda_2$  προς τον δακτύλιο. Την ίδια στιγμή

όλα τα πακέτα στα μήκη κύματος  $\lambda_1$  και  $\lambda_2$  που προωθούνται στον δακτύλιο φαίνονται στο στιγμιότυπο Ε της ίδιας εικόνας. Σημειώνεται ότι είναι προφανές ότι δεδομένης της ύπαρξης μόνο ενός  $1 \times 2$  διακόπτη εντός του WSS, τα πακέτα που είναι γραμμένα στο  $\lambda_1$  προωθούνται όλα στον δακτύλιο.



*Εικόνα 81: (Α-Ε) Στιγμιότυπα των πακέτων που αντιστοιχούν ίδια σημεία (Α)-(Ε) που επισημαίνονται στην Εικόνα 80.*

Η καμπύλη BER ως συνάρτηση της λαμβανόμενης οπτικής ισχύος φαίνεται στην Εικόνα 82. Τα αποτελέσματα δείχνουν πολύ παρόμοια απόδοση με το σενάριο intra-pod, με λειτουργία χωρίς σφάλματα που παρατηρείται για λαμβανόμενες ισχύς μεγαλύτερες από -5,8 dBm. Η χρήση του FPGA ως δέκτη επιτρέπει την εκτίμηση BER σε ένα όριο  $10^{-13}$ .

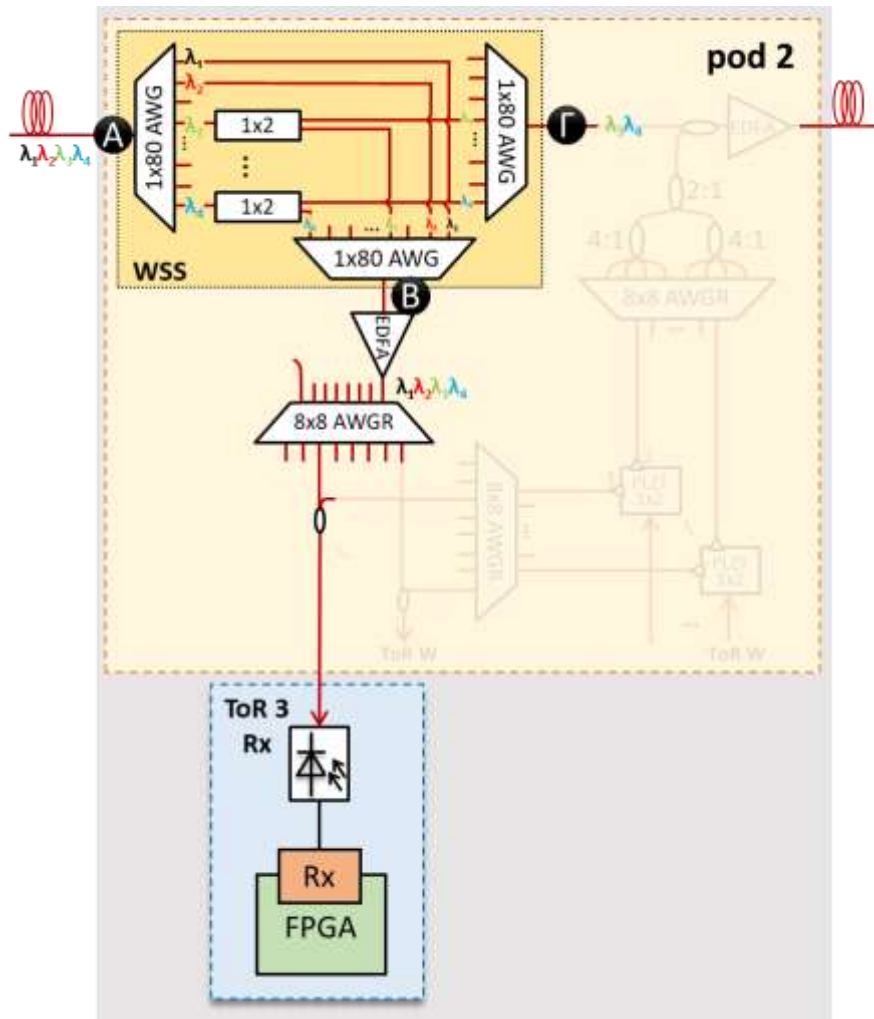


*Εικόνα 82: Το BER ως συνάρτηση της λαμβανόμενης ισχύος για το inter-pod σενάριο*

### 3.3.4 Αξιολόγηση μέσω Inter-pod δρομολόγησης με έμφαση στο WSS

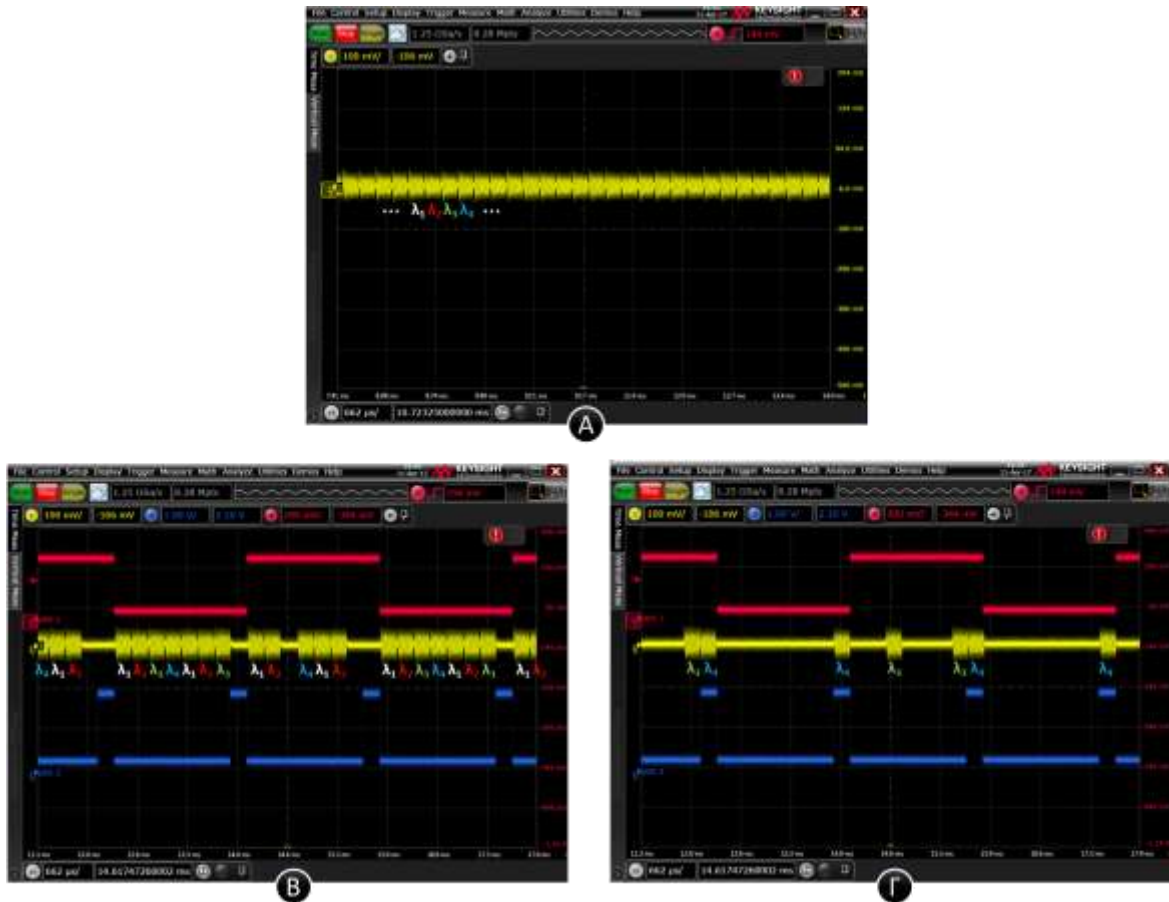
Ένα τελικό σενάριο αξιολόγησης της inter-pod επικοινωνίας πραγματοποιήθηκε ξεχωριστά, εστιάζοντας στη λειτουργικότητα του διακόπτη επιλογής μήκους κύματος WSS. Η Εικόνα 83 δείχνει την τροποποιημένη πειραματική ρύθμιση με έμφαση στην προσέγγιση «αποπολυπλεξία, μεταγωγή και πολυπλεξία» του WSS, όπου σε αυτήν την περίπτωση χρησιμοποιήθηκαν δύο  $1 \times 2$  οπτικοί διακόπτες ώστε να επιτευχθεί δυναμικός έλεγχος δύο μηκών κύματος από το WSS. Μια συνεχής ροή πακέτων που είναι εγγεγραμμένα σε τέσσερα μήκη κύματος ( $\lambda_1=1550,116$  nm,  $\lambda_2=1548,515$  nm,  $\lambda_3=1550,918$  nm,  $\lambda_4=1552,11524$  nm) εισέρχεται στο αριστερό  $1 \times 8$  AWG του WSS, όπως φαίνεται στην Εικόνα 84 (A). Μετά την αποπολυπλεξία, τα  $\lambda_1$  και  $\lambda_2$  συνδέονται στις αντίστοιχες εισόδους του  $8 \times 1$  AWG που βρίσκεται στο κάτω μέρος ενώ τα  $\lambda_3$  και  $\lambda_4$  ελέγχονται από τους δύο διακόπτες  $1 \times 2$ , έτσι ώστε τα αντίστοιχα πλαίσια να μπορούν να κατευθυνθούν μέσα στο pod 2 ή να προωθηθούν στον DWDM δακτύλιο, σύμφωνα με τα σήματα οδήγησης που εφαρμόζονται στους διακόπτες.



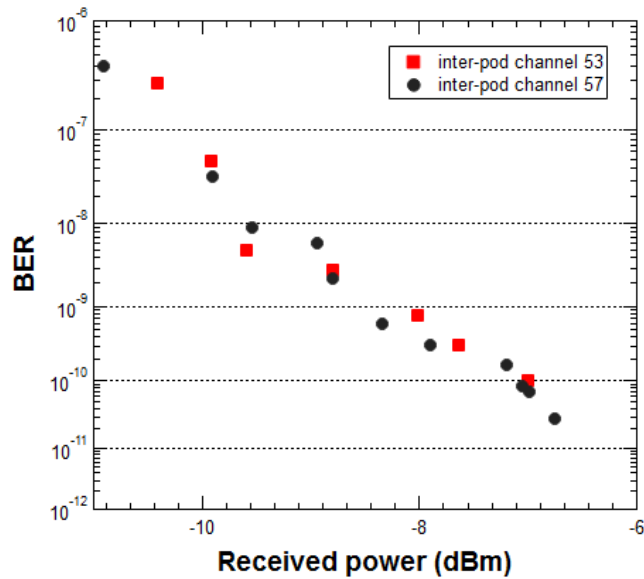


**Εικόνα 83:** Πειραματική διάταξη που χρησιμοποιείται για την αξιολόγηση του σεναρίου επικοινωνίας μεταξύ δυο pod με έμφαση στη λειτουργικότητα του WSS.

Τα στιγμιότυπα (B) και (Γ) της εικόνας 84 απεικονίζουν τις δύο εξόδους του διακόπτη επιλογής μήκους κύματος WSS. Μετά τον παθητικό οπτικό δρομολογητή 8×8 AWGR, τα πακέτα που είναι εγγεγραμμένα στα μήκη κύματος  $\lambda_3$  και  $\lambda_4$  λαμβάνονται ξεχωριστά από τον δέκτη ToR, ο οποίος συνδέεται με τις κατάλληλες εξόδους του AWGR. Τέλος, η Εικόνα 85 δείχνει την καμπύλη BER ως συνάρτηση της λαμβανόμενης οπτικής ισχύος. Πιο συγκεκριμένα για οπτική ισχύ μεγαλύτερη από -6,8dBm δεν παρατηρήθηκαν σφάλματα, αφού το BER είναι καλύτερο από  $3 \cdot 10^{-13}$ .



*Εικόνα 84: Στιγμιότυπα των πακέτων που εισέρχονται και εξέρχονται από το WSS. Τα στιγμιότυπα A, B και Γ αντιπροσωπεύουν τα σημεία που υποδεικνύονται στην πειραματική διάταξη της εικόνας 83.*



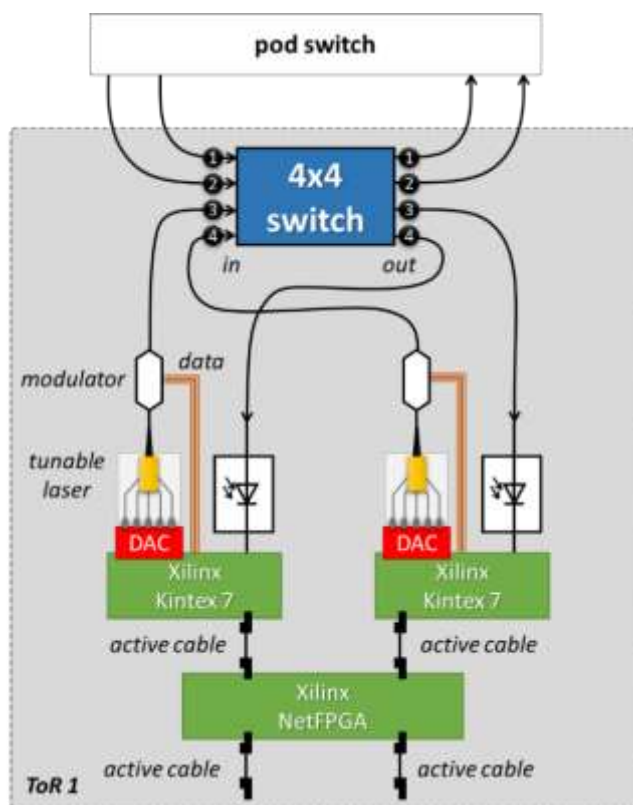
*Εικόνα 85: Το BER ως συνάρτηση της λαμβανόμενης ισχύος για το inter-pod σενάριο με έμφαση στη λειτουργικότητα του WSS*

### 3.4 Ανάπτυξη και πειραματική αξιολόγηση του οπτικού τμήματος του ToR switch με τη χρήση του 4×4 οπτικού διακόπτη PLZT

Στα πειράματα που περιγράφονται σε αυτήν την ενότητα, το κύριο οπτικό στοιχείο που εμπλέκεται είναι ο γρήγορος 4×4 οπτικός διακόπτης PLZT που χαρακτηρίστηκε στην ενότητα 2.3. Ο διακόπτης αυτός είναι υπεύθυνος για την δρομολόγηση εντός του ToR switch (intra-ToR communication) καθώς και για την οπτική επικοινωνία μεταξύ διαφορετικών ToR που βρίσκονται είτε εντός του ίδιου pod (intra-pod communication) ή σε διαφορετικά pod (inter-pod communication). Τα παρακάτω πειράματα σκοπό έχουν να αναδείξουν όλες τις οπτικές διαδρομές (all optical paths) εντός του δικτύου και διεξήχθησαν τηρώντας τις προδιαγραφές της προτεινόμενης αρχιτεκτονικής που περιγράφηκαν με λεπτομέρεια στο ΚΕΦΑΛΑΙΟ 1.

Το οπτικό μέρος του ToR switch όπως περιγράφεται στην ενότητα 1.3.1.2 περιλαμβάνει και έναν οπτικό διακόπτη που επιτρέπει την επικοινωνία με οπτικό τρόπο μεταξύ τερματικών μέσα στο ίδιο ToR (intra-ToR). Σε αυτή την ενότητα παρουσιάζεται η χρήση ενός γρήγορου οπτικού διακόπτη 4×4, θεωρώντας μια τοπολογία που αποτελείται από δύο οπτικά επίπεδα (optical planes), δηλαδή κάθε ToR switch είναι εξοπλισμένο με δύο πομπούς και δύο δέκτες. Ο διακόπτης 4×4 τοποθετείται στην έξοδο κάθε πομπού και πριν από κάθε δέκτη και στην πραγματικότητα εξυπηρετεί δύο κύριες λειτουργίες: από την πλευρά του πομπού δρομολογεί την εξερχόμενη κίνηση από τους ToR είτε ως *intra-ToR* είτε *inter-ToR* ενώ από την άλλη πλευρά διαχειρίζεται την εισερχόμενη κυκλοφορία που προέρχεται από το pod switch και τη δρομολογεί προς τα κατάλληλα τερματικά.

Στην Εικόνα 86 απεικονίζεται μια απλοποιημένη μορφή της αρχιτεκτονικής του ToR switch θεωρώντας δύο επίπεδα οπτικής μεταγωγής και συνεπώς δύο ζεύγη από πομπούς-δέκτες που εξυπηρετούν το rack. Πιο συγκεκριμένα, δύο burst-mode προσαρμοζόμενοι πομποί εντός του ToR switch, που ελέγχονται μέσω δύο ψηφιακών FPGA (Xilinx Kintex KC707) παράγουν οπτικά πακέτα με τα κατάλληλα χαρακτηριστικά. Κάθε πομπός θα μπορεί να στέλνει σε διαφορετικά επίπεδα του δικτύου ανάλογα με το μήκος κύματος που θα εκπέμψει. Επιπλέον, το ToR switch είναι εξοπλισμένο και με ίσο αριθμό φωτοδεκτών, έτσι ώστε να μπορεί να λαμβάνει σήματα από τα διαφορετικά επίπεδα δικτύου.

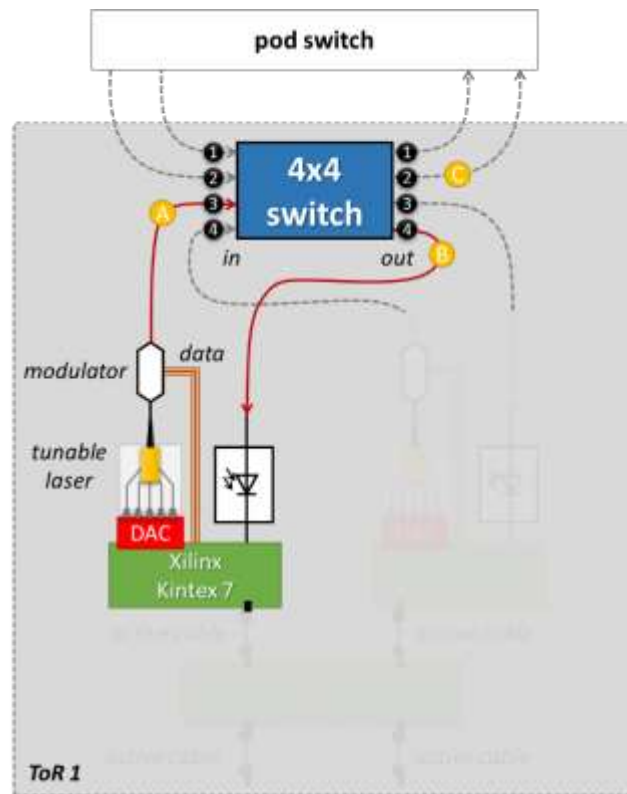


*Εικόνα 86: Επισκόπηση της πειραματικής διάταξης και των κύριων στοιχείων που χρησιμοποιήθηκαν για την ενσωμάτωση του 4×4 οπτικού διακόπτη στο ToR switch.*

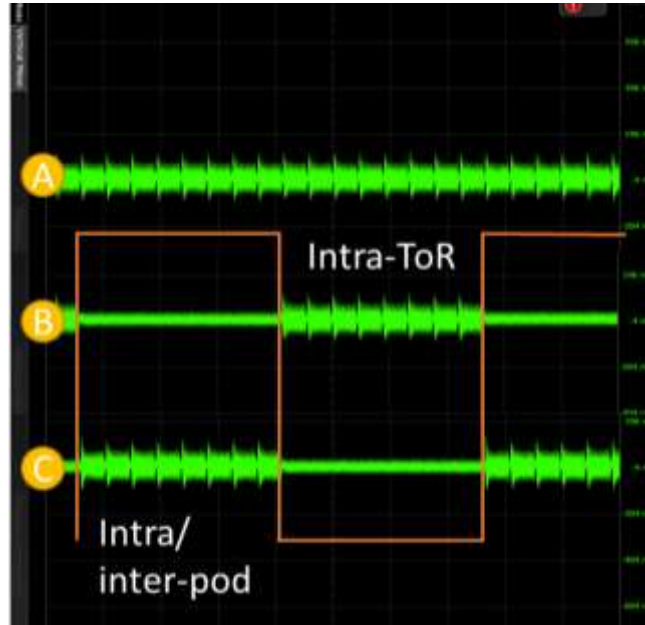
### 3.4.1 Επικοινωνία εντός του ίδιου ToR (Intra-ToR) με εξολοκλήρου οπτικό τρόπο.

Η αξιολόγηση της επιτυχούς επικοινωνίας στον οπτικό τομέα πραγματοποιήθηκε αρχικά μέσω της πειραματικής μελέτης του σεναρίου που αφορά στην επικοινωνία εσωτερικά του ToR. Εδώ να υπενθυμίσουμε ότι η επικοινωνία εσωτερικά του ίδιου ToR σημαίνει ότι χρήστες/υπολογιστές εντός του ίδιου rack επικοινωνούν αποκλειστικά με οπτικό τρόπο μέσω του ToR switch. Όπως φαίνεται στην Εικόνα 87, ο διακόπτης ToR χρησιμεύει τόσο ως πομπός όσο και ως δέκτης καθώς οι κεντρικοί υπολογιστές πηγής και προορισμού συνδέονται με το συγκεκριμένο ToR. Το FPGA δημιουργεί συνεχή ηλεκτρικά πακέτα σε

ρυθμό 10G/s που τροφοδοτούνται στον διαμορφωτή. Ταυτόχρονα, μέσω του μετατροπέα ψηφιακού σήματος σε αναλογικό (Texas Instruments DAC), το FPGA ελέγχει το μήκος κύματος που εκπέμπει το προσαρμοζόμενο λέιζερ. Πρέπει να σημειωθεί ότι στο συγκεκριμένο πείραμα το μήκος κύματος του λέιζερ δεν παίζει κάποιο ρόλο ούτε καθορίζει τον προορισμό, καθώς η διαδρομή δεδομένων επικοινωνίας εντός του ToR δεν περιλαμβάνει στοιχεία επιλογής μήκους κύματος και ως εκ τούτου οι σύνδεσμοι εντός ToR είναι «διαφανείς», δηλαδή αγνωστικιστικοί στο μήκος κύματος.

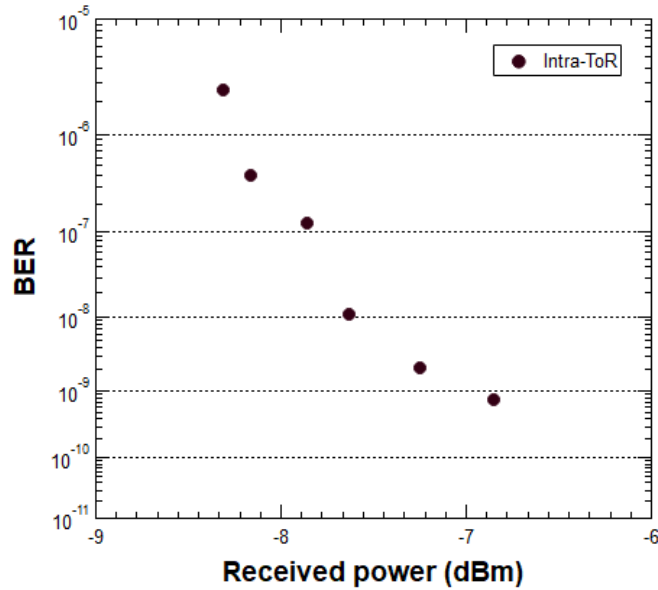


**Εικόνα 87:** Η πειραματική διάταξη που χρησιμοποιήθηκε για την αξιολόγηση του intra-ToR σεναρίου εξολοκλήρου με οπτικό τρόπο.



**Εικόνα 88:** Τα ψηφιακά πακέτα 10G/s όπως αυτά δρομολογούνται μέσω του 4×4 PLZT οπτικού διακόπτη και κατευθύνονται είτε μέσα στο ίδιο ToR (intra-ToR scenario) είτε σε διαφορετικό (intra/inter-pod scenarios). Τα σημεία A,B και C στο σχήμα αντιστοιχούν στα ίδια σημεία στην Εικόνα 88.

Σύμφωνα με την πειραματική διάταξη στην Εικόνα 88, η έξοδος του λέιζερ τροφοδοτείται στον διαμορφωτή και τα διαμορφωμένα οπτικά πακέτα κατευθύνονται προς την είσοδο 3 του 4×4 PLZT οπτικού διακόπτη. Παράλληλα, ο διακόπτης ελέγχεται μέσω του FPGA που παράγει TTL σήματα ελέγχου. Έτσι, η εισερχόμενη στον διακόπτη κίνηση των πακέτων κατευθύνεται προς διαφορετικές εξόδους του ανάλογα με τον τελικό προορισμό τους (προς το ίδιο ή διαφορετικό ToR). Στην υπό εξέταση περίπτωση, όπως απεικονίζεται στην Εικόνα 88, ο διακόπτης κατευθύνει εναλλάξ 8 πακέτα προς την έξοδο 4 (που προορίζονται στον ίδιο ToR) ενώ τα ακόλουθα 8 πακέτα κατευθύνονται προς τη έξοδο 2 που αντιστοιχεί σε κίνηση που προορίζεται σε διαφορετικό ToR. Η δεύτερη περίπτωση (inter-ToR) θα συζητηθεί διεξοδικά σε επόμενες ενότητες, ενώ η αξιολόγηση του σεναρίου intra-ToR πραγματοποιήθηκε με τη λήψη μετρήσεων ρυθμού εσφαλμένων bit (Bit-error-rate) μέσω του FPGA. Οι μετρήσεις αυτές αποτυπώνονται στην Εικόνα 89 υποδεικνύοντας αποδεκτή απόδοση για ένα ευρύ φάσμα λαμβανόμενης οπτικής ισχύος. Πιο συγκεκριμένα, για λαμβανόμενη ισχύ μεγαλύτερη από - 6,67 dBm δεν παρατηρήθηκαν σφάλματα (error-free transmission), που αντιστοιχούν σε BER καλύτερο από  $3 \times 10^{-13}$ .

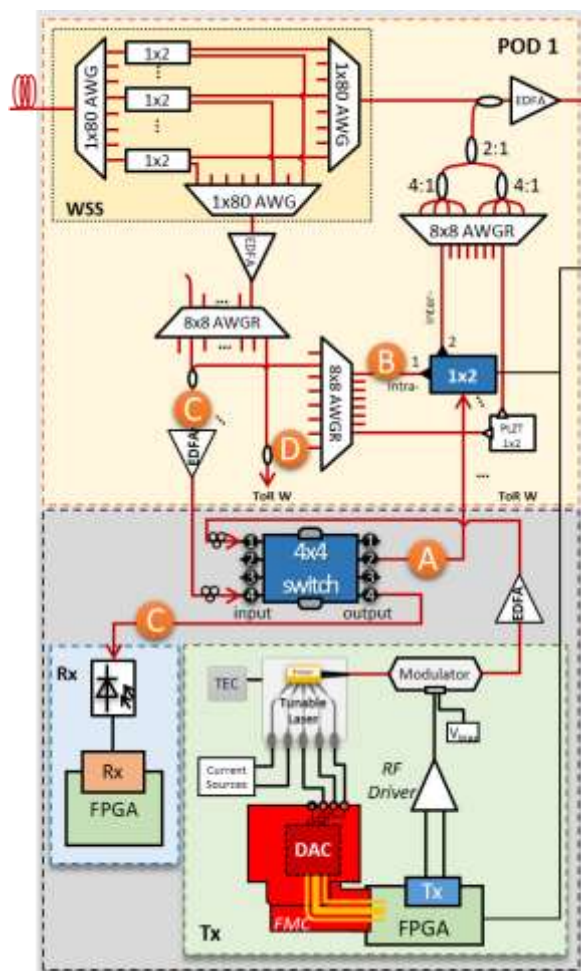


*Εικόνα 89: Ο ρυθμός εσφαλμένων bit (BER) σε σχέση με την λαμβανόμενη οπτική ισχύ για το intra-ToR σενάριο.*

### 3.4.2 Επικοινωνία σε ToR εντός του ίδιου pod με εξολοκλήρου οπτικό τρόπο

Η πειραματική ρύθμιση που απεικονίζεται στην Εικόνα 90 χρησιμοποιήθηκε για την αξιολόγηση της επικοινωνίας μεταξύ υπολογιστών που βρίσκονται εντός του ίδιου pod αλλά είναι συνδεδεμένοι σε διαφορετικούς ToRs (intra-pod σενάριο) με εξολοκλήρου οπτικό τρόπο. Όπως συνέβη και κατά τη διάρκεια του προηγούμενου πειράματος, ο 4×4 PLZT οπτικός διακόπτης τοποθετείται στον ToR μεταγωγέα, ενώ σε αυτήν την περίπτωση ένας οπτικός ενισχυτής ίνας ερβίου (EDFA) χρησιμοποιείται πριν από το διακόπτη προκειμένου να αντισταθμιστεί η απώλεια οπτικής ισχύος. Με την ίδια αρχή λειτουργίας του οπτικού διακόπτη, τα πακέτα που εξέρχονται από την έξοδο 2 και άρα κατευθύνονται σε κάποιον προορισμό εκτός του ToR, εισέρχονται σε έναν νέο 1×2 οπτικό διακόπτη ο οποίος καθορίζει την επικοινωνία εντός του ίδιου pod ή εκτός του pod. (intra/inter-pod switch). Αυτός ο 1×2 οπτικός διακόπτης ελέγχεται από το ToR FPGA, που στην ουσία παράγει και ελέγχει όλο το υποσύστημα. Εν συνεχεία και εφόσον περιγράψουμε μια intra-pod επικοινωνία, χρησιμοποιείται ένα παθητικός οπτικός δρομολογητής (AWGR) 8×8 που κατευθύνει την οπτική κίνηση σε βάση το μήκος κύματος  $\lambda$ .

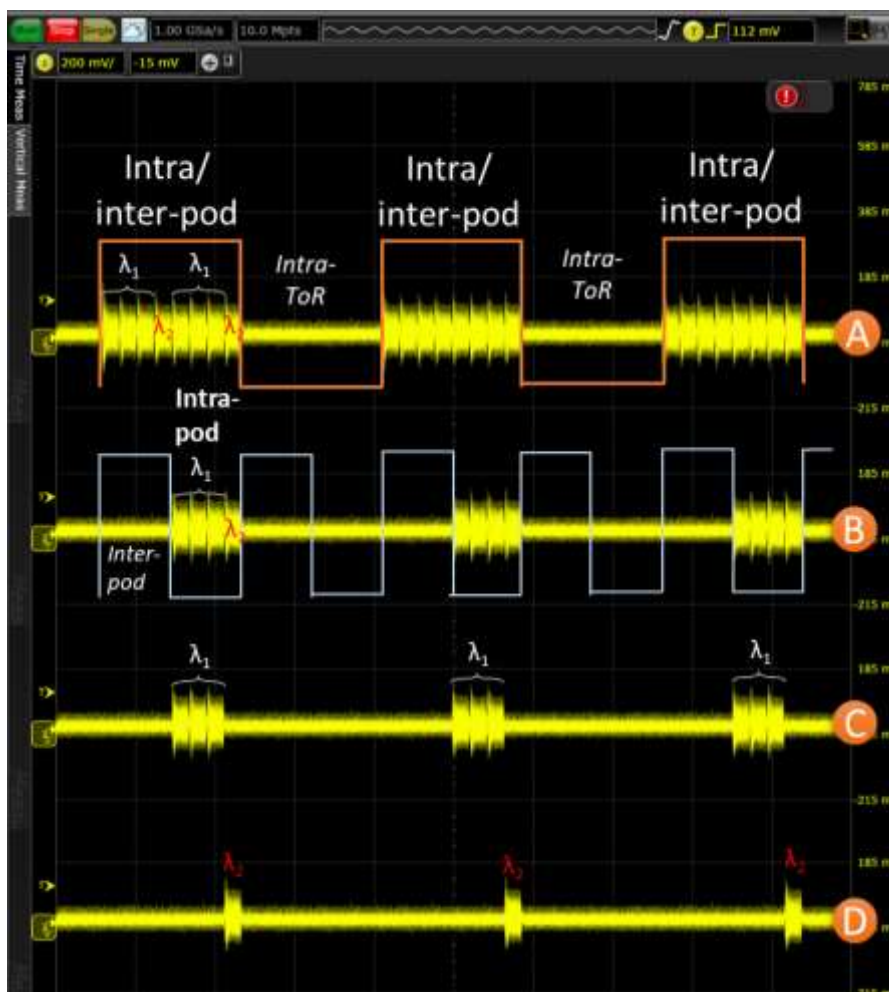
Τα πακέτα με προορισμό μέσα στο ίδιο pod αφού ενισχυθούν μέσω ενός οπτικού ενισχυτή EDFA προωθούνται και πάλι μέσω του 4×4 PLZT οπτικού διακόπτη. Αξίζει να σημειωθεί ότι χρησιμοποιήθηκε ο ίδιος διακόπτης 4×4 και στη μεριά του δέκτη αλλά και σε αυτήν του πομπού δεδομένης της έλλειψης δεύτερου διακόπτη τη στιγμή της διεξαγωγής του πειράματος.



**Εικόνα 90:** Πειραματική διάταξη για την επικοινωνία ToR εντός του ίδιου pod με εξολοκλήρου οπτικό τρόπο. Τα σημεία A,B,C και D που διακρίνονται στο σχήμα αντιστοιχίζονται με αυτά στην Εικόνα 91 και δείχνει τον διαχωρισμό των πακέτων ανά μήκος κύματος και φυσικά ανάλογα με τη δρομολόγηση των οπτικών διακοπών.

Ο προσαρμοζόμενος πομπός εκπέμπει εναλλάξ σε δύο μήκη κύματος:  $\lambda_1=1546,91$  nm και  $\lambda_2=1551,72$  nm. Πιο συγκεκριμένα η εναλλαγή των μηκών κύματος γίνεται με τέτοιο τρόπο ώστε τρία πακέτα να εγγράφονται στο  $\lambda_1$  ενώ ένα πακέτο να εγγράφεται στο  $\lambda_2$ . Αυτή η ακολουθία πακέτων και μηκών κύματος επαναλαμβάνεται συνεχώς, ενώ τα πακέτα που αντιστοιχούν σε κάθε μήκος κύματος φαίνονται χαρακτηριστικά στην Εικόνα 91.

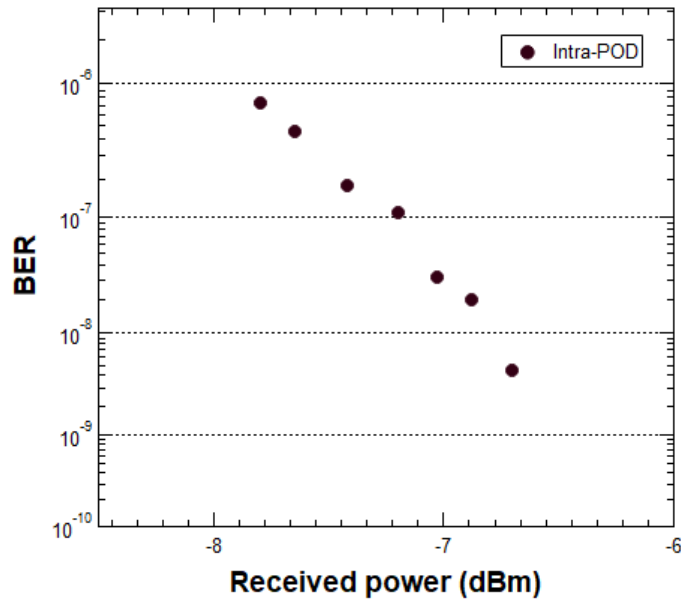




*Εικόνα 91: Απεικόνιση των πακέτων που δρομολογούνται μέσω του 4×4 PLZT διακόπτη αλλά και του 1×2 διακόπτη που καθορίζει intra/inter-pod επικοινωνία. Τα σημεία A,B,C και D που διακρίνονται στο σχήμα αντιστοιχίζονται με αυτά στην Εικόνα 90 και δείχνει τον διαχωρισμό των πακέτων ανά μήκος κύματος και φυσικά ανάλογα με τη δρομολόγηση των οπτικών διακοπών.*

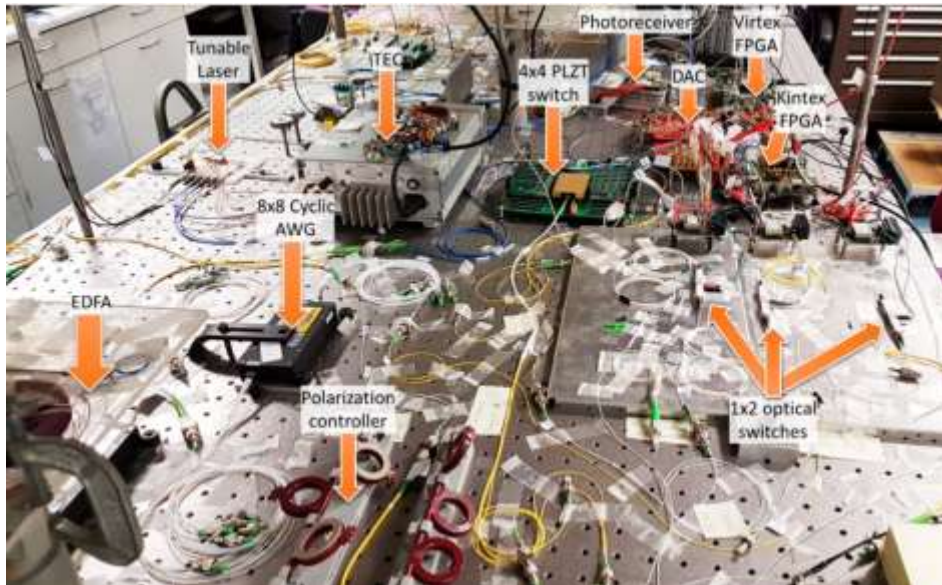
Όπως και στο προηγούμενο σενάριο, ο 4×4 PLZT οπτικός διακόπτης καθορίζει την intra-ToR από την inter-ToR επικοινωνία. Έτσι, έχει ρυθμιστεί ώστε να εναλλάσσει την κίνηση προς τις εξόδους 1 και 2 κάθε 8 πακέτα (Εικόνα 91 A), ενώ ο διακόπτης 1×2 εκτρέπει την κίνηση προς το ίδιο (intra-pod) ή άλλο pod (inter-pod) κάθε 4 πακέτα (Εικόνα 91 B). Τέλος, όλη η κίνηση που τροφοδοτείται στον διακόπτη προορισμού 4×4 κατευθύνεται στον δέκτη ToR. Όπως φαίνεται στην Εικόνα 91 D που αντιπροσωπεύει τη λαμβανόμενη κίνηση σε αυτό το σενάριο και με τις παραπάνω ελεγκτικές ρυθμίσεις στους διακόπτες, τότε στην περίοδο των 16 πακέτων, μόνο το ένα από αυτά «έφτασε» στον δέκτη του ToR. Επιπλέον

αξιολόγηση της απόδοσης της διαδρομής επικοινωνίας εντός του pod έγινε μέσω μετρήσεων BER. Η Εικόνα 92 απεικονίζει την αντίστοιχη καμπύλη BER ως συνάρτηση της λαμβανόμενης οπτικής ισχύος. Για λαμβανόμενη ισχύ μεγαλύτερη από -6,64 dBm δεν παρατηρήθηκαν σφάλματα, που αντιστοιχούν σε BER καλύτερο από  $3 \cdot 10^{-13}$ .



*Εικόνα 92: Μέτρηση BER σε σχέση με την λαμβανόμενη οπτική ισχύ για το σενάριο που αφορά επικοινωνία εντός του ίδιου pod με εξολοκλήρου οπτικό τρόπο.*

Τέλος αξίζει να παραθέσουμε και μια φωτογραφία από την πειραματική διάταξη, όπως φαίνεται στην Εικόνα 93. Η συγκεκριμένη διάταξη που παρουσιάζει μεγάλο βαθμό πολυπλοκότητας δεδομένου ότι αποτελείται από πολλά μέρη, αναπτύχθηκε στο Εργαστήριο Φωτονικών Επικοινωνιών και χρησιμοποιήθηκε για την αξιολόγηση των intra- και inter-pod σεναρίων επικοινωνίας. Όλα τα οπτικά και ηλεκτρικά υποσυστήματα συνδυάστηκαν υπό διαφορετικές ρυθμίσεις και διατάξεις με λεπτομερή και προσεκτική διασύνδεση αφού μόνο έτσι θα είχαμε επιτυχή ολοκλήρωση των προαναφερθέντων πειραμάτων.



*Εικόνα 93: Φωτογραφία από την πειραματική διάταξη όπως αυτή αναπτύχθηκε στο Εργαστήριο Φωτονικών Επικοινωνιών (ΕΦΕ) στο ΕΜΠ.*

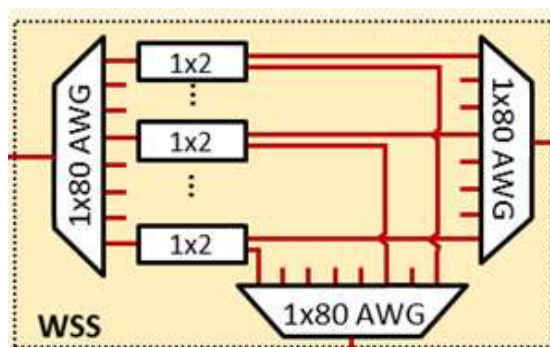
### 3.4.3 Επικοινωνία μεταξύ ToRs σε διαφορετικά pod με εξολοκλήρου οπτικό τρόπο

Ως τελευταίο βήμα για την αξιολόγηση της ποιότητας της εξολοκλήρου οπτικής επικοινωνίας, σχεδιάστηκε και υλοποιήθηκε πειραματική διάταξη για το σενάριο inter-pod, ολοκληρώνοντας έτσι την αξιολόγηση για όλες τις πιθανές διαδρομές επικοινωνίας σε μια αρχιτεκτονική όπως αυτή που προτείνεται στα πλαίσια της διατριβής.

Στην πειραματική διάταξη που περιγράφεται σε αυτήν την υποενότητα και απεικονίζεται στην Εικόνα 95, η πλευρά του πομπού είναι πανομοιότυπη με αυτήν στις προηγούμενες περιπτώσεις. Ωστόσο, σε αυτήν την περίπτωση η κίνηση κατευθύνεται από ένα pod προς ένα διαφορετικό pod ύστερα από κατάλληλη ρύθμιση του 1x2 διακόπτη, που όπως αναφέρθηκε και προηγουμένως είναι ο διακόπτης καθορισμού της intra- από της inter-pod επικοινωνίας.

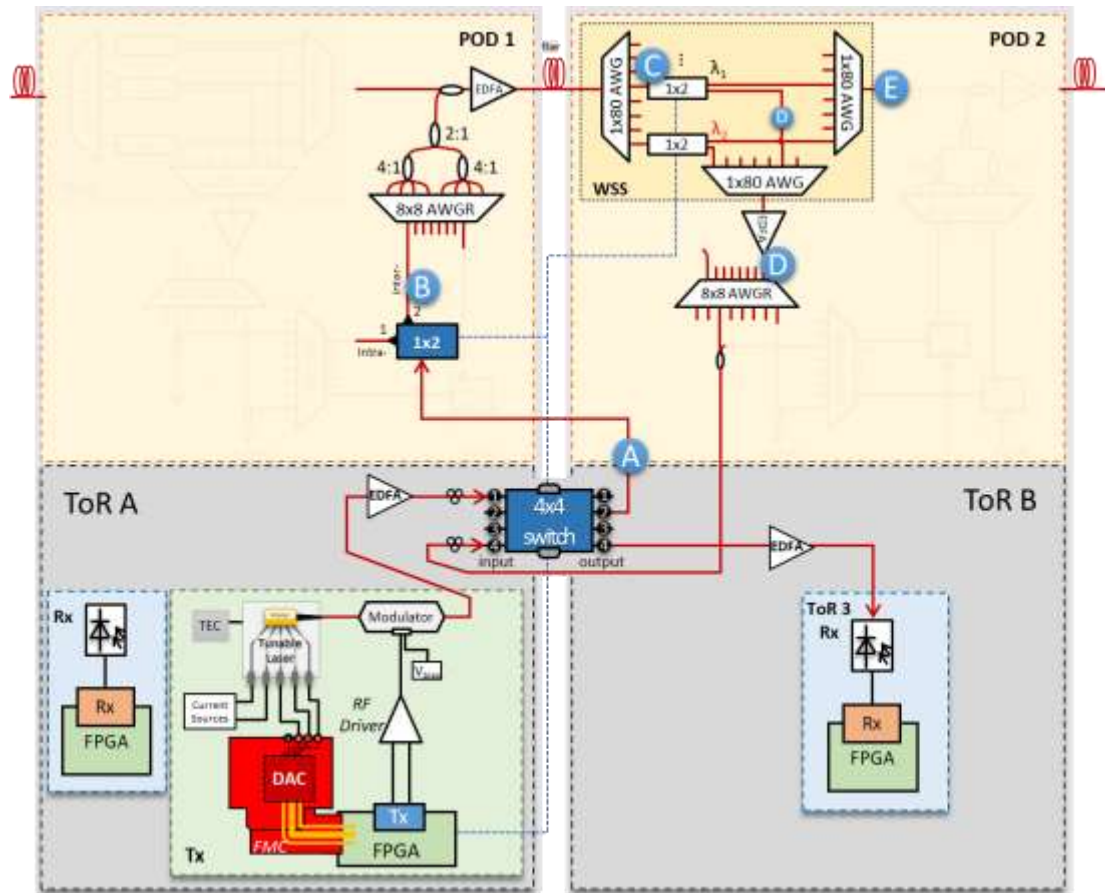
Έπειτα η δρομολόγηση γίνεται μέσα από τον διακόπτη επιλογής μήκους κύματος (Wavelength Selective Switch) στο pod switch του pod προορισμού. Πιο συγκεκριμένα για τον διακόπτη επιλογής μήκους κύματος (εν συντομία WSS), δεν χρησιμοποιήθηκε κάποιος εμπορικά διαθέσιμος διακόπτης αλλά έγινε μια υλοποίηση χρησιμοποιώντας επιμέρους οπτικά μέρη, ώστε να μπορεί να ανταπεξέλθει στις απαιτήσεις της προτεινόμενης αρχιτεκτονικής. Αναλυτικός χαρακτηρισμός του συγκεκριμένου οπτικού υποσυστήματος έχει γίνει αλλά δεν αφορά την συγκεκριμένη διατριβή. Η υλοποίηση που ακολουθήθηκε για το WSS βασίζεται στην «αποπολυπλεξία, μεταγωγή και πολυπλεξία» (Εικόνα 94), όπου οπτικά φίλτρα (AWG) και επιμέρους 1x2 οπτικοί διακόπτες ανά μήκος κύματος συνδυάζονται ώστε να επιτρέπουν τον δυναμικό έλεγχο (μέσω σημάτων ελέγχου από

FPGA) διαφορετικών μηκών κύματος μέσα στο WSS και με αυτόν τον τρόπο να δρομολογούν την οπτική κίνηση ανά  $\lambda$  είτε μέσα στο rod είτε να την προωθούν στον ξανά στον DWDM δακτύλιο.



**Εικόνα 94:** Ο διακόπτης επιλογής μήκους κύματος (Wavelength Selective Switch) που υλοποιήθηκε βασισμένος στην προσέγγιση «αποπολυπλεξία, μεταγωγή και πολυπλεξία» για την δυναμική επιλογή μήκους κύματος.

Όσον αφορά στο πείραμα που περιγράφεται παρακάτω, χρησιμοποιήθηκαν δύο 1x2 οπτικοί διακόπτες που ο καθένας αφορά στην διαχείριση της κατεύθυνσης συγκεκριμένου μήκους κύματος. Όπως φαίνεται και στην Εικόνα 95 το WSS κατευθύνει το  $\lambda_1$  στο εσωτερικό του rod, ενώ το  $\lambda_2$  προωθείται στον οπτικό δακτύλιο, προφανώς γιατί ο προορισμός του βρίσκεται σε κάποιο άλλο rod. Έπειτα από την χρήση παθητικών οπτικών φίλτρων δρομολόγησης (AWGRs), οπτικοί ενισχυτές χρησιμοποιούνται ώστε να διατηρηθεί η οπτική ισχύς σε αποδεκτά επίπεδα για λήψη χωρίς σφάλματα. Τέλος, η αρχιτεκτονική του δέκτη του ToR switch είναι πανομοιότυπη με αυτή που χρησιμοποιείται στο σενάριο intra-rod, και μοιράζεται την ίδια μονάδα διακόπτη 4x4 PLZT με τον πομπό.



**Εικόνα 95:** Πειραματική διάταξη για την εξολοκλήρου οπτική επικοινωνία ToRs που βρίσκονται σε διαφορετικά rod. Τα σημεία A,B,C,D και E που διακρίνονται στο σχήμα αντιστοιχίζουν με αυτά στην Εικόνα 96 και δείχνει τον διαχωρισμό των πακέτων ανά μήκος κύματος και φυσικά ανάλογα με τη δρομολόγηση των οπτικών διακοπών.

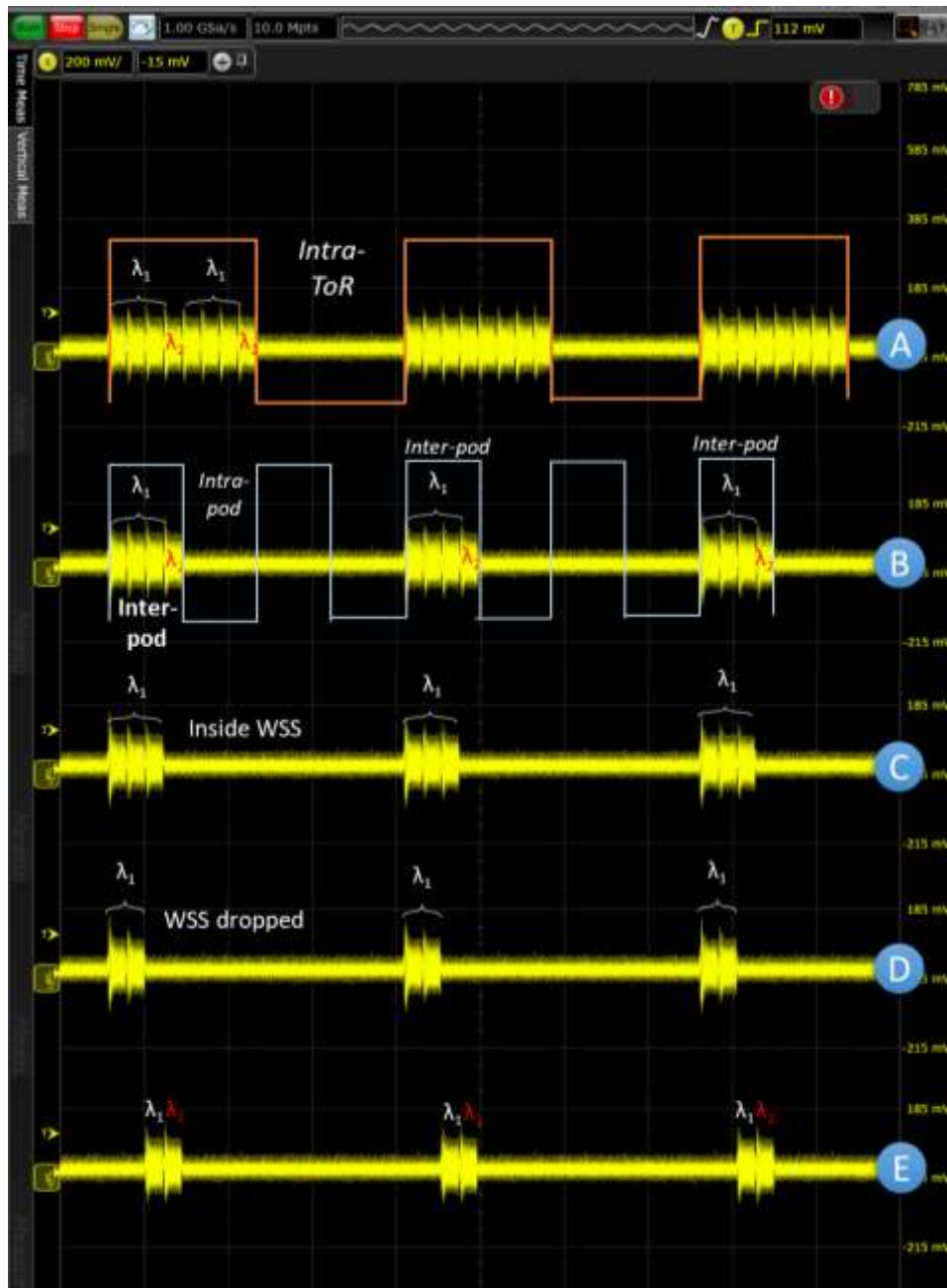
Η Εικόνα 96 – Α απεικονίζει τα πακέτα που παράγονται από τον πομπό του ToR switch που βρίσκεται στο rod 1 και η κίνησή τους διαχωρίζεται από τον 4x4 PLZT οπτικό διακόπτη σε intra- και inter-ToR. Τα πακέτα εκείνα που κατευθύνονται εκτός του ToR, δηλαδή έχουν προορισμό ένα άλλο ToR είτε εντός του rod 1 είτε σε κάποιο άλλο δρομολογούνται από τον οπτικό διακόπτη 1x2 του rod 1, που βρίσκεται στην θέση B στην Εικόνα 95, ενώ η κίνηση που προωθείται προς τον οπτικό δακτύλιο, δηλαδή σε μια inter-rod διαδρομή, φαίνεται στην Εικόνα 96 – B. Εκεί επίσης φαίνεται, ότι με μια περιοδικότητα τρία πακέτα γραμμένα στο  $\lambda_1$  και ένα πακέτο γραμμένο στο  $\lambda_2$  έχουν προορισμό κάποιο ToR που βρίσκεται σε άλλο rod.

Στο επόμενο στιγμιότυπο που φαίνεται στην Εικόνα 96 – C διακρίνουμε τα τρία πακέτα στο  $\lambda_1$  που βρίσκονται μέσα στο WSS και πλέον έχουν αποπολυπλεχτεί μέσω του οπτικού φίλτρου και πρόκειται να δρομολογηθούν από τον αντίστοιχο 1x2 διακόπτη είτε μέσα στο

rod 2 είτε να προωθηθούν σε άλλο rod μέσω του οπτικού δακτυλίου. Ακολουθώντας την διαδρομή των πακέτων του  $\lambda_1$  παρατηρούμε ότι δύο από τα τρία πακέτα παραμένουν μέσω του WSS στο rod 1, αφού προορίζονται σε κάποιο ToR μέσα σε αυτό (Εικόνα 96 – D).

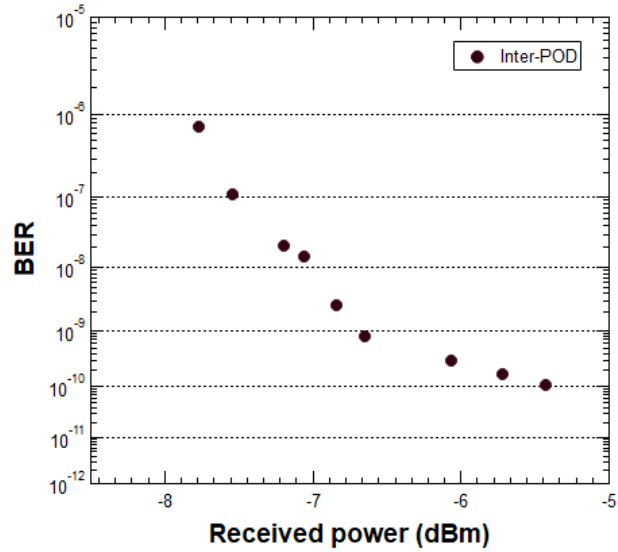
Αντίθετα, όπως φαίνεται στην Εικόνα 96 – E το πακέτο του  $\lambda_1$  που απόμεινε, πολυπλέχτηκε με το μοναδικό πακέτο του  $\lambda_2$  που ακολούθησε πανόμοια διαδικασία και μαζί προωθούνται στο επόμενο rod, ώστε να φτάσουν στον προορισμό τους. Φυσικά, όλη παραπάνω διαδικασία έγινε με αμιγώς οπτικό τρόπο με ότι αυτό συνεπάγεται για τον συγχρονισμό και την πολυπλοκότητα.

Όπως συνέβη και στις προηγούμενες πειραματικές περιπτώσεις που παρατέθηκαν στις προηγούμενες υποενότητες, έτσι και εδώ πραγματοποιήθηκαν μετρήσεις BER από το FPGA στη μεριά του δέκτη έτσι ώστε να αξιολογηθεί η απόδοση ολόκληρης της inter-rod διαδρομής.



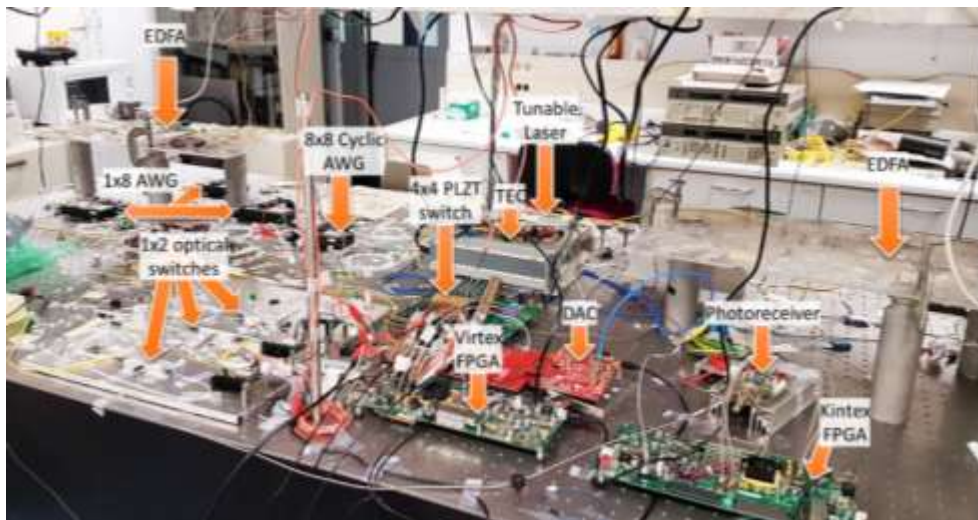
**Εικόνα 96:** Απεικόνιση των πακέτων που δρομολογούνται μέσω του  $4 \times 4$  PLZT διακόπτη, του  $1 \times 2$  διακόπτη που καθορίζει intra/inter-pod επικοινωνία και φυσικά μέσω του WSS. Τα σημεία A,B,C,D και E που διακρίνονται στο σχήμα αντιστοιχίζονται με αυτά στην Εικόνα 95 και δείχνει τον διαχωρισμό των πακέτων ανά μήκος κύματος και φυσικά ανάλογα με τη δρομολόγηση των οπτικών διακοπών.

Στην Εικόνα 97 φαίνεται η καμπύλη BER ως συνάρτηση της λαμβανόμενης οπτικής ισχύος, ενώ για οπτική ισχύ μεγαλύτερη από  $-4,82$  dBm δεν παρατηρήθηκαν σφάλματα (BER καλύτερο από  $3 \cdot 10^{-13}$ ).



**Εικόνα 97:** Μέτρηση BER σε σχέση με την λαμβανόμενη οπτική ισχύ για το σενάριο που αφορά την οπτική επικοινωνία μεταξύ υπολογιστών που βρίσκονται σε διαφορετικά pod.

Η Εικόνα 98 δείχνει φωτογραφία της πειραματικής διάταξης που χρησιμοποιήθηκε για την επικύρωση του inter-pod σεναρίου. Όπως και στις προηγούμενες περιπτώσεις, η συνεπής συμπεριφορά όλων των οπτικών και ηλεκτρικών εξαρτημάτων ήταν απαραίτητη προϋπόθεση για την επιτυχή ολοκλήρωση του πειράματος.



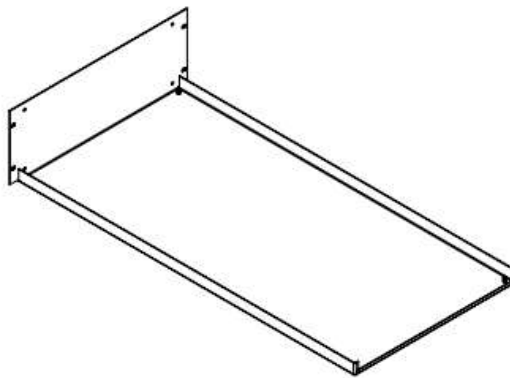
**Εικόνα 98:** Φωτογραφία από την πειραματική διάταξη για την intra-pod επικοινωνία όπως αυτή αναπτύχθηκε στο Εργαστήριο Φωτονικών Επικοινωνιών (ΕΦΕ) στο ΕΜΠ.



### 3.5 Ενοποίηση των υποσυστημάτων σε μια ενιαία οπτική δομή

Σε αυτή την ενότητα, παρουσιάζεται φωτογραφικό υλικό και σχεδιαγράμματα από τη διαδικασία συναρμολόγησης και διασύνδεσης, όλων των επιμέρους υποσυστημάτων σε ένα ενιαίο οπτικό σύστημα. Η διαδικασία αυτή είχε ως τελικό στόχο την δημιουργία ενός μικρού πρότυπου οπτικού datacenter στο Εργαστήριο Φωτονικών Επικοινωνιών της Σχολής Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών ΕΜΠ. Αυτό το μικρό πρότυπο datacenter (που η δομή και η λειτουργικότητά του περιγράφεται λεπτομερώς στο επόμενο κεφάλαιο), αποτελείται από συνολικά 4 ToRs και 2 rod switches που διασυνδέονται μέσω 2 οπτικών επιπέδων. Όπως έχει αναλυθεί στις προηγούμενες ενότητες του Κεφαλαίου 3, η υλοποίηση καθενός από τα οπτικά υποσυστήματα είναι μια πολύ απαιτητική και δύσκολη διαδικασία δεδομένου ότι πρέπει να συναρμολογηθούν και να ελεγχθούν πολυάριθμα εξαρτήματα συνδεδεμένα είτε μέσω οπτικών ινών είτε μέσω ηλεκτρικών καλωδίων.

Πρώτο βήμα στη διαδικασία κατασκευής του πρότυπου datacenter ήταν ο σχεδιασμός και η κατασκευή 9 μεταλλικών ραφιών (διάστασης 4U - Εικόνα 99) που το καθένα θα περιέχει ένα διακριτό οπτικό υποσύστημα, ενώ όλα μαζί θα τοποθετηθούν σε ένα μεταλλικό rack (Εικόνα 100). Το σχέδιο ενός τέτοιου ραφιού φαίνεται στην Εικόνα 99, ενώ θα πρέπει να επισημανθεί η αντιστατική ιδιότητά του για την προστασία των ευαίσθητων φωτονικών και ηλεκτρονικών συστημάτων έναντι του στατικού ηλεκτρισμού.

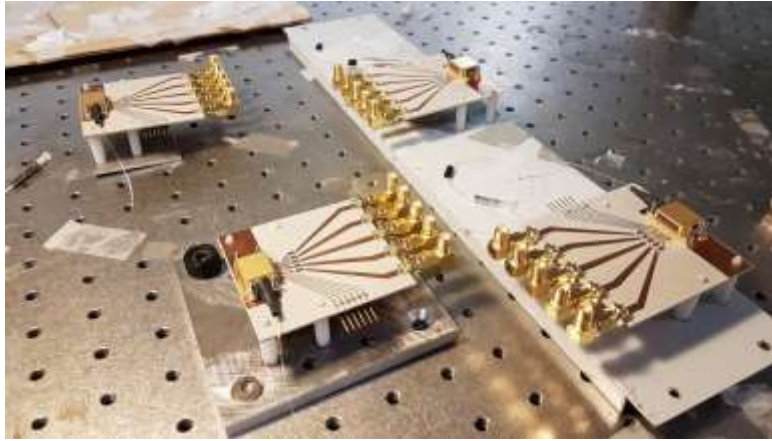


*Εικόνα 99: Σχεδιασμός του μεταλλικού αντιστατικού ραφιού που θα περιέχει τα οπτικά υποσυστήματα στο πρότυπο datacenter*



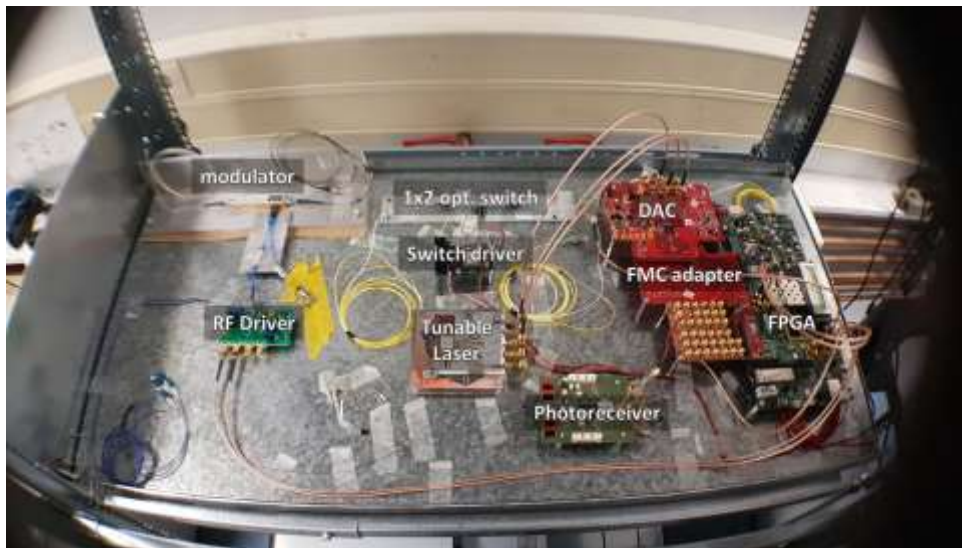
*Εικόνα 100: Το rack του πρότυπου datacenter, που φιλοξενεί τα ράφια*

Το επόμενο βήμα ήταν ο χαρακτηρισμός και η βαθμονόμηση όλων των όμοιων συστημάτων, ώστε να λειτουργούν όλα με τον ίδιο τρόπο. Για παράδειγμα, όσον αφορά τα λέιζερ των προσαρμοζόμενων πομπών στα ToRs, χρησιμοποιήθηκαν έξι ίδια λέιζερ τα οποία φαίνονται στην Εικόνα 101 και έπρεπε να ρυθμιστούν αναλόγως, ώστε πιθανές παραλλαγές στα PCB boards να μην επηρεάζουν το μήκος κύματος που τελικά εκπέμπεται. Αυτή η διαδικασία ακολουθήθηκε για όλα τα οπτικά και ηλεκτρικά συστήματα όπως είναι οι οπτικοί διακόπτες, οι RF ενισχυτές, οι οπτικοί διαμορφωτές και τα οπτικά φίλτρα δρομολόγησης.



*Εικόνα 101: Τα όμοια λέιζερ των πομπών κατά τη διαδικασία χαρακτηρισμού τους.*

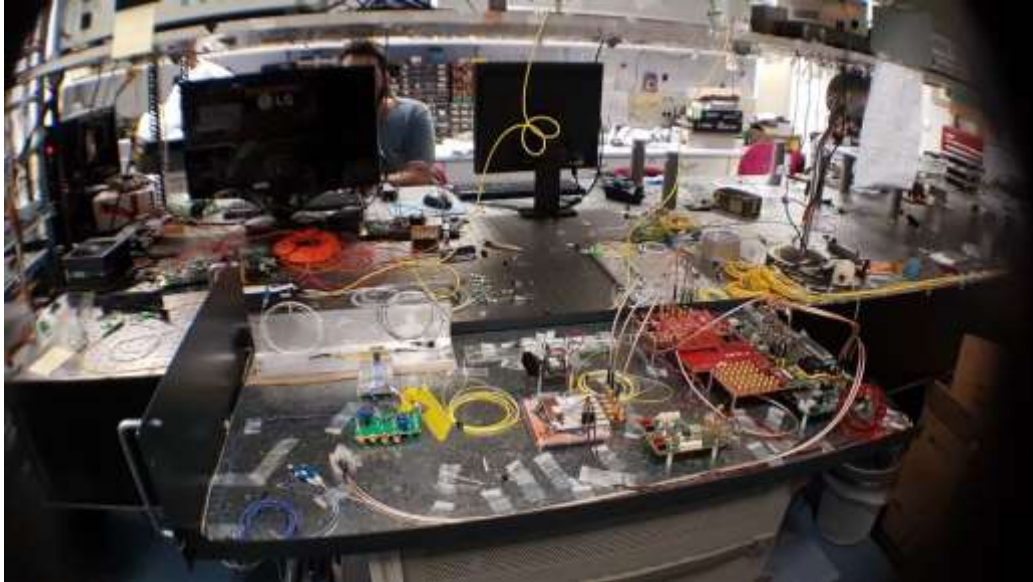
Κατά τη διαδικασία συναρμολόγησης, επίσης, έπρεπε να ληφθούν υπόψη σημαντικά ζητήματα που αφορούσαν σε αποκλίσεις μεταξύ των όμοιων υποσυστημάτων, οι οποίες οφείλονταν αφενός στις μικρές αποκλίσεις λόγω κατασκευής, και αφετέρου στο γεγονός ότι όλο το σύστημα στήθηκε με το χέρι με ότι αυτό συνεπάγεται στα διάφορα μήκη των οπτικών ινών, τις κολλήσεις των ηλεκτρικών διεπαφών και τις διασυνδέσεις όλων αυτών μεταξύ τους.



*Εικόνα 102: Ένα από τα ράφια που φιλοξενεί όλα τα υποσυστήματα του ToR switch*

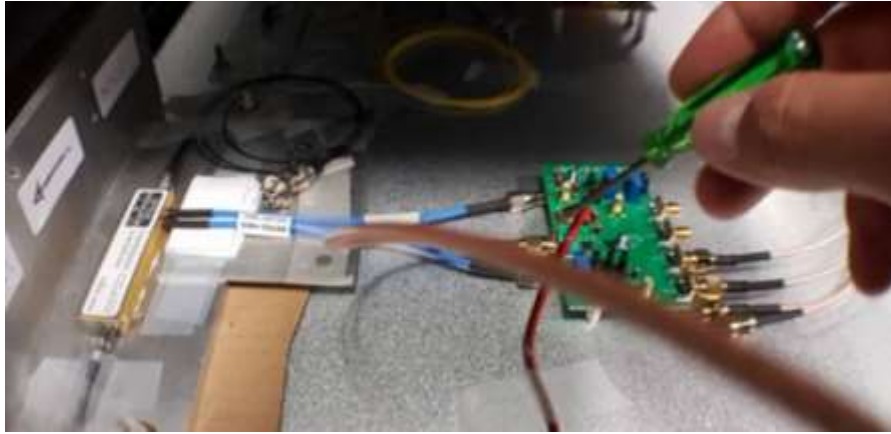
Στην Εικόνα 102 φαίνεται ένα από τα ράφια που περιέχει «μισό» ToR switch. Για να γίνει πιο κατανοητό αυτό, όλοι οι ToR switches χωρίστηκαν στη μέση, ώστε να μπορούν να εξυπηρετούν δύο διαφορετικά οπτικά επίπεδα (optical planes). Επιπλέον, τόσο τα υποσυστήματα πομπού όσο και ο δέκτης χωρούσαν σε ένα ενιαίο ράφι. Με αυτόν τον τρόπο, διευκολύνθηκε η διαδικασία χαρακτηρισμού. Είναι σαφές ότι αυτή η τμηματική

προσέγγιση που ακολουθεί την κλιμακούμενη λογική της προτεινόμενης αρχιτεκτονικής, επιτρέπει την επέκταση του συνολικού συστήματος προσθέτοντας περισσότερα διακριτά ToR υποσυστήματα.



*Εικόνα 103: Φωτογραφία κατά τη διαδικασία δοκιμών ενός ToR switch.*

Φωτογραφίες από τη διαδικασία χαρακτηρισμού και βαθμονόμησης διαφορετικών ηλεκτρικών και οπτικών συστημάτων απεικονίζονται στις εικόνες 103 και 104. Αρχικά, τα όμοια συστήματα βαθμονομήθηκαν μεμονωμένα, με στόχο την επίτευξη όσο το δυνατόν πιο παρόμοιας απόδοσης μεταξύ τους. Η πειραματική διάταξη σχεδιάστηκε με τέτοιο τρόπο ώστε η επιλογή των καλωδίων, των ινών και η τοποθέτηση των συστημάτων σε κάθε ράφι να ακολουθεί μια συνεπή μεθοδολογία. Η ίδια διαδικασία δοκιμών και βαθμονόμησης ακολουθήθηκε για κάθε διάταξη ToR και pod switch, γεγονός που διευκόλυνε τη βήμα προς βήμα ενσωμάτωσή τους στο τελικό rack του πρότυπου datacenter (Εικόνα 105).



*Εικόνα 104: Βαθμονόμηση του RF driver έτσι ώστε να επιτυγχάνεται η ίδια απόδοση μεταξύ όλων των πομπών ToR.*

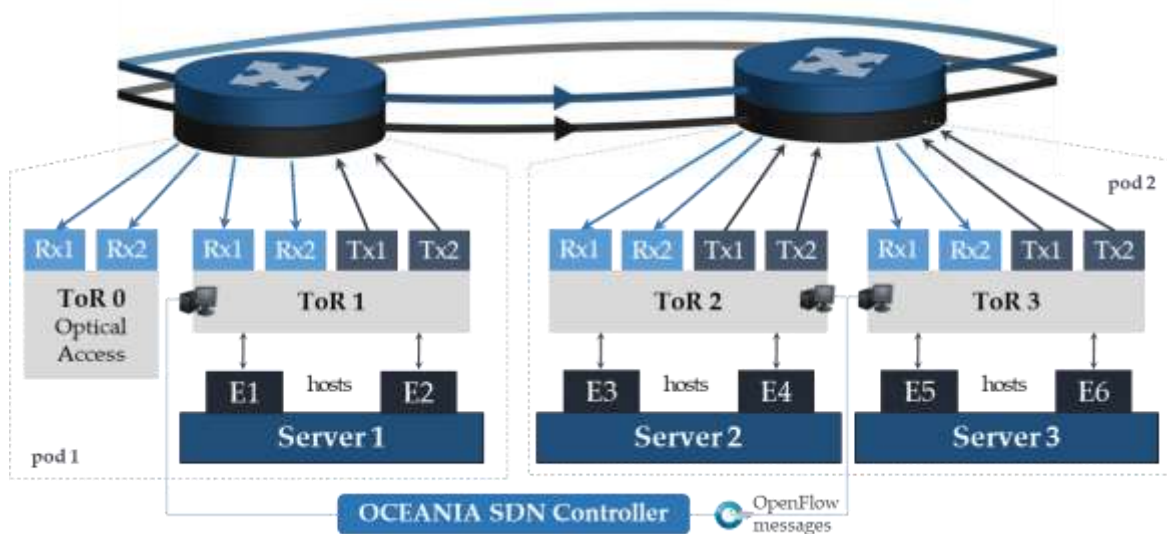


*Εικόνα 105: Το ολοκληρωμένο rack του πρότυπου mini-datacenter, που περιέχει όλα τα υποσυστήματα*

# ΚΕΦΑΛΑΙΟ 4

## Το πρωτότυπο μίνι datacenter στο ΕΦΕ

Σε αυτό το κεφάλαιο, παρουσιάζεται το πρωτότυπο μίνι datacenter (mini-DCN) που κατασκευάστηκε βήμα-βήμα στο Εργαστήριο Φωτονικών Επικοινωνιών (ΕΦΕ) του Εθνικού Μετσόβιου Πολυτεχνείου και που υλοποιεί σε μικρή κλίμακα την προτεινόμενη υβριδική αρχιτεκτονική. Σκοπός του πρωτότυπου mini-DCN ήταν να αναδειχθεί ο αρμονικός συντονισμός όλων των επιμέρους δομικών μερών του από το φυσικό στρώμα (physical layer) μέχρι το επίπεδο δικτύου (network orchestration), ενώ χρησιμοποιήθηκε για να γίνει απόδειξη του ότι η προτεινόμενη αρχιτεκτονική μπορεί αν δουλέψει αποτελεσματικά. Οι φωτογραφίες του πρωτοτύπου κατά την περίοδο προετοιμασίας του παρουσιάστηκαν στην αναλυτικά στην ενότητα 3.5.



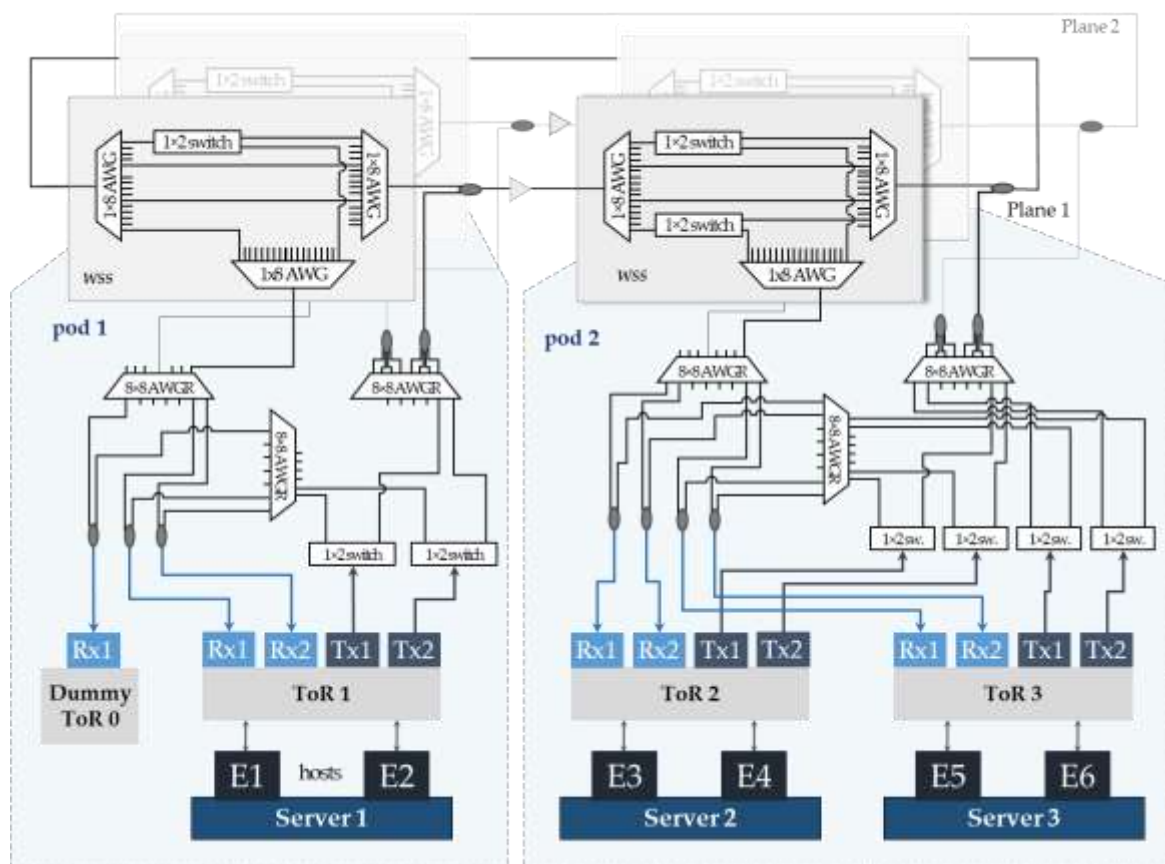
*Εικόνα 106: Το διάγραμμα της αρχιτεκτονικής δομής του πρωτότυπου datacenter*

### 4.1 Η αρχιτεκτονική δομή και η κλίμακα του πρωτότυπου datacenter

Το διάγραμμα της αρχιτεκτονικής δομής του πρωτότυπου datacenter απεικονίζεται στο σχήμα της Εικόνας 106. Με τη συγκεκριμένη δομή προσομοιώνονται δύο οπτικά επίπεδα (optical planes) με δύο pod στα πρότυπα της προτεινόμενης αρχιτεκτονικής. Πιο συγκεκριμένα, το pod 1 φιλοξενεί αφενός ένα διακομιστή (Server 1) με δύο ανεξάρτητες διεπαφές δικτύου Ethernet 10 Gb/s (E1 και E2), που επικοινωνούν με το υπόλοιπο δίκτυο

μέσω του ToR 1 και αφετέρου ένα εικονικό ToR (ToR 0) που χρησιμοποιείται για σκοπούς παρακολούθησης του δικτύου. Από την άλλη πλευρά, το pod 2 φιλοξενεί δύο διακομιστές (Server 2 και Server 3), ο καθένας από τους οποίους επικοινωνεί μέσω των ToR 2 (E3 και E4) και ToR 3 (E5 και E6) αντίστοιχα.

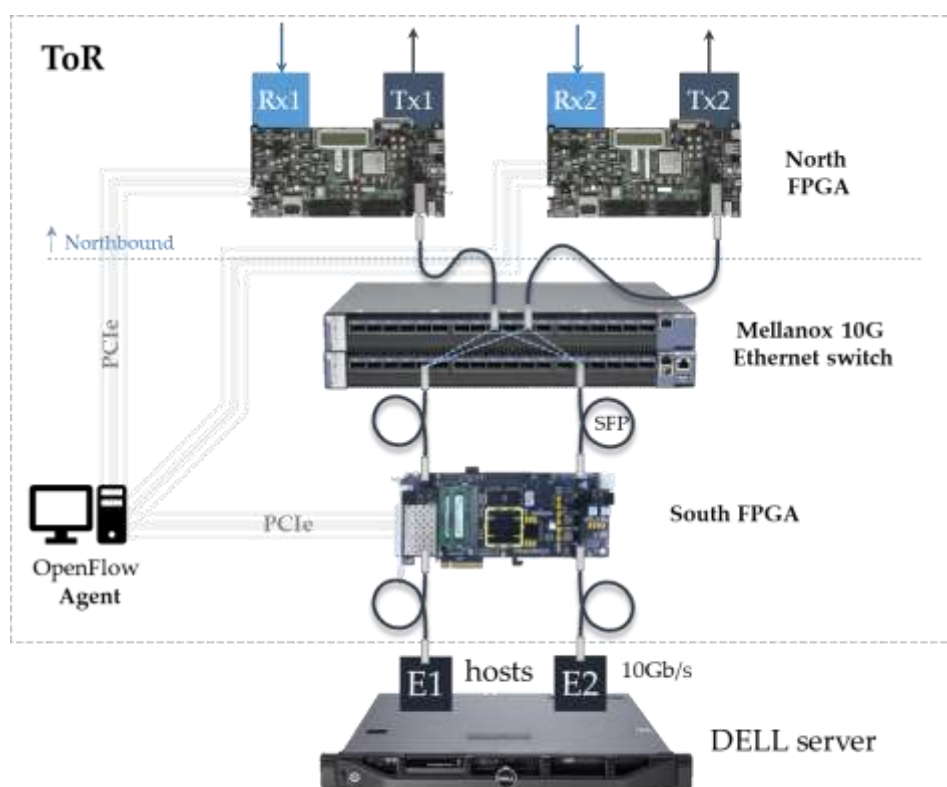
Όπως προαναφέρθηκε και στο εισαγωγικό του κεφαλαίου, το πρωτότυπο μίνι datacenter υλοποιήθηκε σύμφωνα με τις προτεινόμενες αρχιτεκτονικές αρχές αλλά σε διαστάσεις εργαστηριακής κλίμακας για τους σκοπούς της επίδειξης του πρωτοτύπου. Πιο συγκεκριμένα καθένας από τους ToR switches είναι εξοπλισμένος με δύο προσαρμοζόμενους πομπούς (Tx) και ίσο αριθμό φωτοδεκτών (Rx), όπως απεικονίζεται στην Εικόνα 107. Έτσι λοιπόν, το πρωτότυπο datacenter αποτελείται από 3 πλήρως εξοπλισμένους ToR switches, δηλαδή έξι υποσυστήματα πομπών και δεκτών, που αναπτύχθηκαν και δοκιμάστηκαν στα προηγούμενα κεφάλαια. Τέλος, ένας μερικώς λειτουργικός δέκτης συνδέθηκε στο «εικονικό» ToR 0, το οποίο βρίσκεται στο pod 1 και χρησιμοποιήθηκε, όπως προαναφέρθηκε για σκοπούς παρακολούθησης (monitoring).



**Εικόνα 107:** Η διάταξη όλων των υποσυστημάτων που χρησιμοποιήθηκαν κατά την επίδειξη του πρωτότυπου datacenter



Επιπλέον, κάθε rod switch χρησιμοποιεί δύο διακόπτες επιλογής μήκους κύματος (WSS) που βασίζονται στην προσέγγιση «Demultiplex, switch and multiplex» (βλ. υποενότητα 2.4.2), έναν για κάθε οπτικό επίπεδο (optical plane). Φυσικά, πολλαπλά παθητικά οπτικά στοιχεία φιλτραρίσματος (8×8 Arrayed Waveguide Grating Router AWGR) και οπτικοί συζεύκτες συνδυάζονται προκειμένου να επιτευχθούν όλα τα πιθανά σενάρια επικοινωνίας μεταξύ των κεντρικών υπολογιστών (servers), με βασικό εργαλείο δρομολόγησης την εναλλαγή του μήκους κύματος ανάλογα με τον προορισμό. Επιπλέον, χρησιμοποιήθηκαν οπτικοί ενισχυτές (EDFA) για την ενίσχυση των οπτικών σημάτων σε όλες τις οπτικές διαδρομές μεταξύ των pods.



**Εικόνα 108:** Λεπτομερής σχηματική απεικόνιση των συνδέσεων από το κατώτερο επίπεδο του server στο υψηλότερο επίπεδο του πομπού και του δέκτη μέσα σε κάθε ToR.

Η Εικόνα 108 δείχνει μια λεπτομερή σχηματική απεικόνιση των εσωτερικών συνδέσεων μεταξύ του ToR switch και των κεντρικών υπολογιστών (servers). Ξεκινώντας την περιγραφή από κάτω προς τα πάνω, τρεις διακομιστές (Εμπορικά διαθέσιμοι της εταιρείας DELL) χρησιμοποιήθηκαν ως κεντρικοί υπολογιστές μέσα στα rack, που λειτουργούσαν ως ζώνες καινοτομίας (innovation zones). Καθένας από τους διακομιστές μπορεί και δημιουργεί κίνηση Ethernet 10 Gb/s ενώ συνδέεται με το South FPGA (S-FPGA) μέσω δύο διεπαφών SFP 10 Gbps. Οι διεπαφές είναι ανεξάρτητες και ως εκ τούτου μπορούν να θεωρηθούν ως ξεχωριστοί κεντρικοί υπολογιστές. Πράγματι, ο Server 1 που βρίσκεται στο

pod 1, υπό τη διαχείριση του ToR1, είναι συνδεδεμένος στο S-FPGA μέσω των διασυνδέσεων Ethernet E1 και E2. Κατόπιν, η κίνηση προωθείται στον ηλεκτρικό μεταγωγέα Ethernet (Ethernet switch της εταιρείας MELLANOX), ο οποίος στέλνει τα Ethernet πλαίσια στο διαθέσιμο North FPGA (N-FPGA) για μετάδοση. Τα N-FPGA ελέγχουν τον προσαρμοζόμενο πομπό Tx και δημιουργούν τα πακέτα της προτεινόμενης αρχιτεκτονικής (διάρκεια πακέτου: 200μs, συμπεριλαμβανομένου χρόνου προστασίας 10μs). Στη μεριά του πομπού Tx, τα ηλεκτρικά πακέτα, που περιέχουν μεταξύ άλλων και τα πλαίσια Ethernet, ενισχύονται από έναν RF driver και στη συνέχεια εισάγονται σε έναν Mach-Zehnder οπτικό διαμορφωτή. Το προσαρμοζόμενο λέιζερ ελέγχεται επίσης πλήρως από το N-FPGA και εκπέμπει το επιθυμητό μήκος κύματος. Η λήψη των πακέτων ακολουθεί την αντίθετη κατακόρυφη διαδικασία και διαδρομή.

Όσον αφορά στο φυσικό επίπεδο, η επικοινωνία μεταξύ των διακομιστών στο προτεινόμενο δίκτυο επιτυγχάνεται με τη χρήση διαφορετικών μηκών κύματος για τη δρομολόγηση των πακέτων, αξιοποιώντας το συνδυασμένο αποτέλεσμα των γρήγορων λέιζερ και των παθητικών οπτικών δρομολογητών AWGR. Από την άλλη πλευρά, όσον αφορά το στρώμα δικτύου/πρωτοκόλλου, σε κάθε διεπαφή Ethernet εκχωρούνται διαφορετικές IP διευθύνσεις. Ο Πίνακας 11 δείχνει τα μήκη κύματος και τις διευθύνσεις IP που αντιστοιχούν σε κάθε διεπαφή ανάλογα με το pod και το ToR στο οποίο βρίσκονται.

Όσον αφορά το επίπεδο ελέγχου, ο ελεγκτής SDN και ο ενορχηστρωτής του πρωτότυπου μίνι datacenter εκτελούνται σε ένα απομακρυσμένο υπολογιστή που είναι συνδεδεμένος μέσω τοπικού δικτύου με ταχύτητες 1 Gbps. Ο ελεγκτής SDN συνδέεται με τους ανάλογους SDN agents, οι οποίοι εκτελούνται στους υπολογιστές που ελέγχουν όλα τα FPGA μέσω της διεπαφής PCIeexpress. Κατά την επίδειξη του πρωτοτύπου, οι οδηγίες που προέρχονται από το επίπεδο ελέγχου διαμορφώνονται, μεταφέρονται και εφαρμόζονται από τα FPGA σε πραγματικό χρόνο.

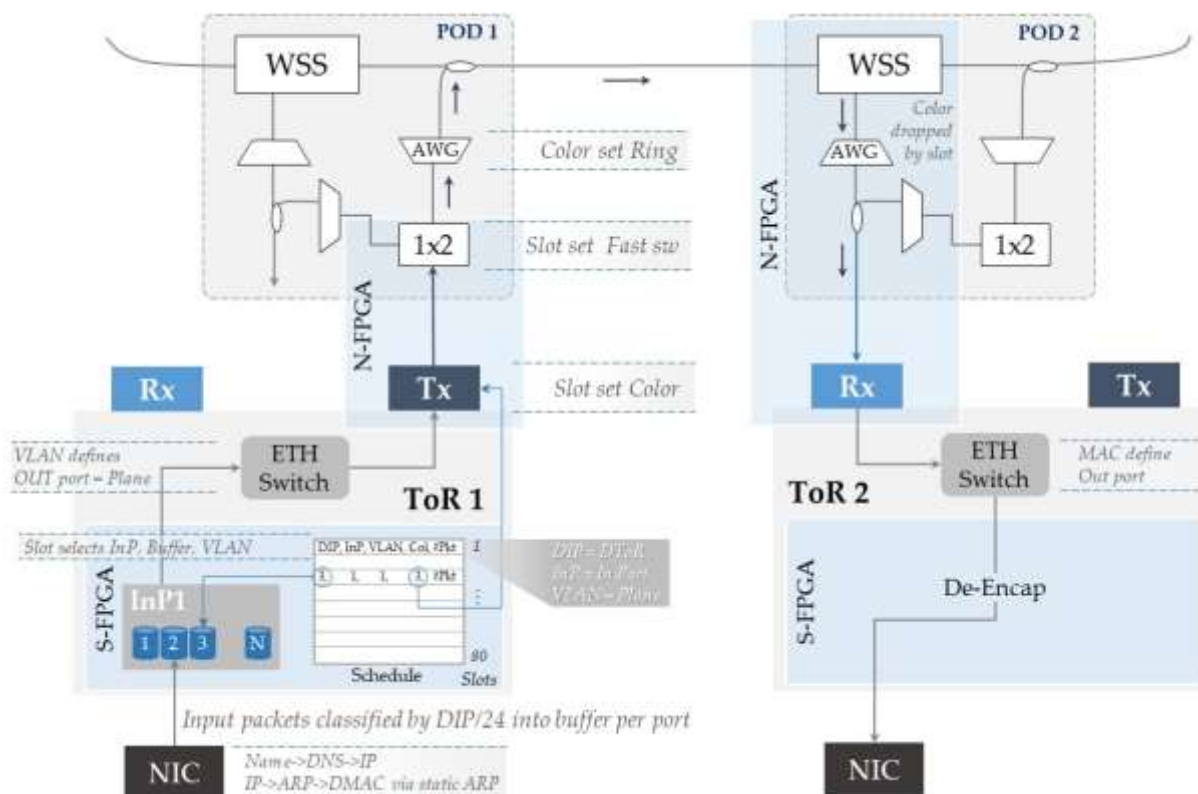
**Πίνακας 11: Μήκη κύματος και IP διευθύνσεις που αντιστοιχούν σε κάθε server σύμφωνα με το pod και το ToR**

ToR	Pod	Μήκος Κύματος (nm)	Server	Δ/νση IP
ToR 1	1	1546.917	E1	10.1.1.1
			E2	10.1.1.129
ToR 2	2	1547.715	E3	10.2.2.1
			E4	10.2.2.129
ToR 3	2	1548.515	E5	10.2.3.1
			E6	10.2.3.129

ToR 4	1	1549.315	Dummy ToR
-------	---	----------	-----------

## 4.2 Μια «ημέρα» στη ζωή ενός πακέτου

Η υποενότητα αυτή δίνει μια περιγραφή της διαδρομής ενός πακέτου μέσα στο προτεινόμενο δίκτυο, με στόχο την καλύτερη κατανόηση της μετάδοσης δεδομένων από άκρο σε άκρο. Κάθε τελικός υπολογιστής διαθέτει μια στατική διεύθυνση IP (Πρωτόκολλο Διαδικτύου) χρησιμοποιώντας DHCP (Πρωτόκολλο διαμόρφωσης δυναμικού κεντρικού υπολογιστή) ή OpenFlow. Κάθε rack είναι ένα υποδίκτυο IP, επομένως οι διευθύνσεις IP όλων των κεντρικών υπολογιστών στο rack ξεκινούν με το ίδιο πρόθεμα (χρησιμοποιούμε πρόθεμα 24-bit). Η διεύθυνση κεντρικού υπολογιστή εντός του υποδικτύου IP είναι το ευρετήριο του NIC (Network Interface Card) εντός του ToR.



Εικόνα 109: Μια «ημέρα» στη ζωή ενός πακέτου

Τα πλαίσια Ethernet μεταδίδονται από το NIC/κεντρικό υπολογιστή στις θύρες εισόδου του ToR (ToR1 στην Εικόνα 109) μέσω του S-FPGA. Αυτά τα πλαίσια φέρουν τη διεύθυνση DMAC (Destination Media Access Control) του κεντρικού υπολογιστή προορισμού. Το

Static ARP (Address Resolution Protocol) μπορεί να χρησιμοποιηθεί για την αντιστοίχιση της διεύθυνσης IP του κεντρικού υπολογιστή προορισμού με την αντίστοιχη διεύθυνση MAC. Οι στατικοί πίνακες ARP θα πρέπει να παρέχονται από τον SDN ελεγκτή. Τα πλαίσια που αποστέλλονται από τους κεντρικούς υπολογιστές φτάνουν στο ToR και αποθηκεύονται σε buffer ταξινομημένα σε ζεύγη θύρας εισόδου και ToR προορισμού. Αυτά τα αποθηκευμένα στον buffer πλαίσια για μια συγκεκριμένη θύρα εισόδου φαίνονται στην Εικόνα 109 μέσα στο πλαίσιο InP1. Το χρονοδιάγραμμα που παρέχεται από τον ελεγκτή SDN και εκτελείται από τον ToR switch, εκτελείται ανά χρονοθυρίδα (slot after slot). Οι οδηγίες που προέρχονται από τον ελεγκτή SDN για κάθε slot υπαγορεύουν αφενός ποιο ζεύγος του buffer [Θύρα Εισόδου, ToR Προορισμού] να σταλεί την συγκεκριμένη χρονική στιγμή, και αφετέρου το μήκος κύματος που θα χρησιμοποιηθεί και μέσω ποιου επιπέδου (plane) θα δρομολογηθεί το πακέτο. Εάν υπάρχουν πακέτα σε αυτό το buffer, αποστέλλονται στον Ethernet μεταγωγέα του ToR αφού επισημανθούν με μια ετικέτα VLAN που ταιριάζει με το επίπεδο προορισμού (destination plane) που ορίζεται από το χρονοδιάγραμμα. Ο Ethernet μεταγωγέας του ToR που καθορίζει την κυκλοφορία μέσα στο ίδιο ToR (intra-ToR) έχει ρυθμιστεί από πριν στατικά μέσω εντολών OpenFlow. Κάθε ετικέτα VLAN αντιστοιχίζεται σε διαφορετικό «North Port» συνδεδεμένο σε pod switch. Με αυτόν τον τρόπο τα εισερχόμενα πλαίσια με ετικέτα VLAN στον Ethernet μεταγωγέα προωθούνται στο σωστό plane.

Όπως το Ethernet πλαίσιο κατευθύνεται πλέον προς το προσαρμοζόμενο λέιζερ που θα καθορίσει το pod προορισμού και το plane, τα πλαίσια Ethernet αφαιρούνται από την ετικέτα VLAN και ενσωματώνονται πλέον σε ένα πακέτο σύμφωνα με τις προδιαγραφές της προτεινόμενης αρχιτεκτονικής. Επίσης ο pod switch ελέγχεται επίσης από έναν OpenFlow agent και εκτελεί το προγραμματισμένο πρόγραμμά του. Σε κάθε μια χρονοθυρίδα ορίζεται και η κατάσταση του γρήγορου διακόπτη καθορισμού αλλά και του WSS (για κάθε μήκος κύματος). Στην τρέχουσα χρονοθυρίδα, το pod 1 γνωρίζει εάν το τρέχον πακέτο αφορά κίνηση Inter-pod ή Intra-pod. Στην περίπτωση κυκλοφορίας Inter-pod, το πακέτο δρομολογείται μέσω του 1×2 οπτικού διακόπτη καθορισμού και του AWGR προς τον οπτικό δακτύλιο.

Στην Εικόνα 109 απεικονίζεται ένα σενάριο inter-pod επικοινωνίας, ενώ το pod 2 έχει εντολή να «κρατήσει» το μήκος κύματος που μεταφέρει τα δεδομένα από το ToR 1 στην τρέχουσα χρονοθυρίδα. Τα πλαίσια που φθάνουν αφαιρούνται από την κεφαλίδα του πακέτου πριν φτάσουν στο διακόπτη Ethernet και έπειτα στον προορισμό τους που είναι το ToR 2. Ο Ethernet μεταγωγέας στο ToR 2 έχει διαμορφωθεί κι αυτός με στατικό τρόπο μέσω OpenFlow για να προωθήσει την κυκλοφορία εντός ToR 2 χρησιμοποιώντας τη διεύθυνση DMAC που είναι ενσωματωμένη στην κεφαλίδα του Ethernet πλαισίου. Με αυτόν τον τρόπο τα πλαίσια Ethernet προωθούνται ανάλογα με τη διεύθυνση DMAC. Προαιρετικά, η προώθηση των πλαισίων θα μπορούσε να έχει πραγματοποιηθεί

χρησιμοποιώντας τη διεύθυνση IP του NIC προορισμού. Οι ίδιοι κανόνες προώθησης του μεταγωγέα Ethernet επιτρέπουν την προώθηση της intra-ToR κίνησης εντός ToR1.

### **4.3 Σενάρια end-to-end επικοινωνίας και αποτελέσματα σε πραγματικό χρόνο**

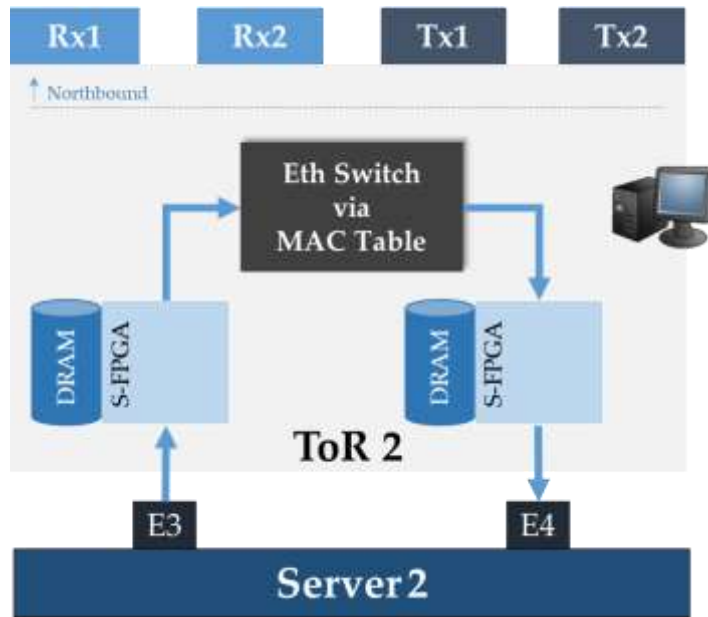
Σε αυτή την ενότητα, παρουσιάζονται τα αποτελέσματα της λειτουργίας του πρωτότυπου μίνι datacenter από άκρο σε άκρο (end-to-end) και σε πραγματικό χρόνο (real-time). Πιο συγκεκριμένα, τα σενάρια επικοινωνίας μέσα στο πρωτότυπο mini-DCN στην υποενότητα 4.3.1, ενώ μια διασύνδεση του πρωτοτύπου της Αθήνας με ένα άλλο datacenter στην Πίζα της Ιταλίας μέσω διαδικτύου παρουσιάζεται στην υποενότητα 4.3.2.

#### **4.3.1 Σενάρια Επικοινωνίας εντός του πρωτότυπου mini-DCN (Intra-DCN)**

Τα σενάρια επικοινωνίας intra-DCN αφορούν σενάρια που λαμβάνουν χώρα εντός του πρωτότυπου mini-DCN. Έχοντας σαν οδηγό το σχηματικό της δομής του mini-DCN που φαίνεται στην Εικόνα 106, τα σενάρια κατηγοριοποιούνται ως εξής: Σενάρια επικοινωνίας Intra-ToR, Intra-pod και Inter-pod, ανάλογα με την τοποθεσία των κεντρικών υπολογιστών που επικοινωνούν εντός του mini-DCN.

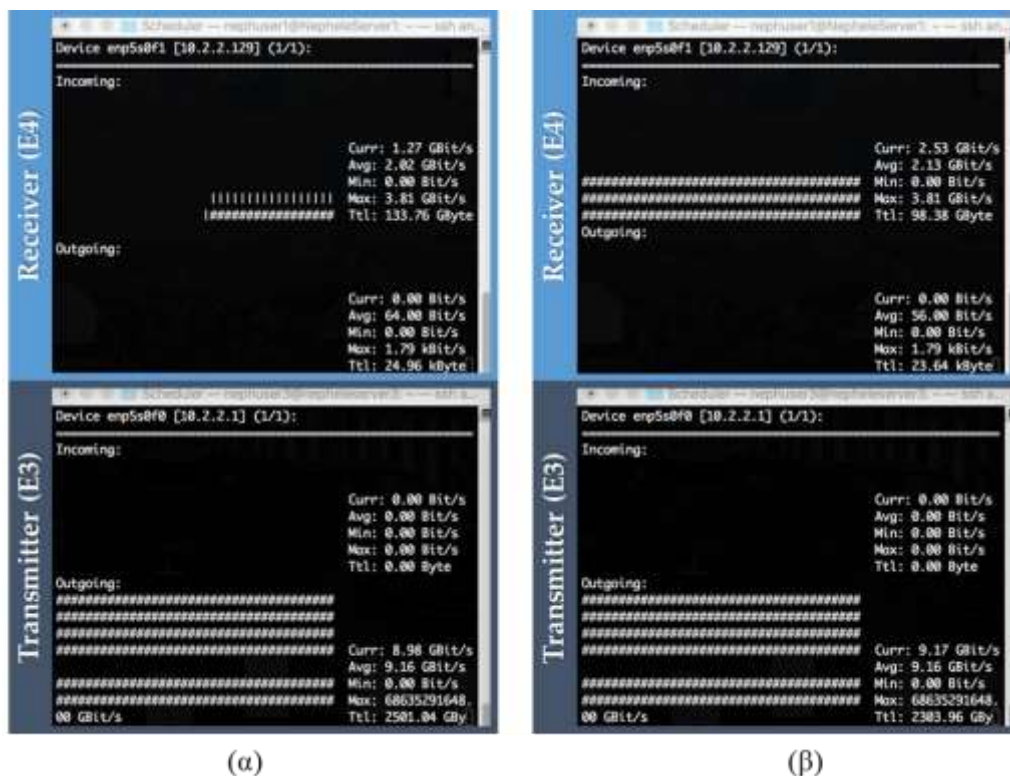
##### **4.3.1.1 Σενάριο Intra-ToR**

Το πρώτο σενάριο που εξετάστηκε ήταν η περίπτωση επικοινωνίας εντός του ίδιου ToR. Σε αυτό το σενάριο, η κίνηση Ethernet παραμένει στο ίδιο rack και ως αποτέλεσμα δεν εμπλέκεται το οπτικό μέρος του ToR switch. Η Εικόνα 110 δείχνει το διάγραμμα της διάταξης Intra-ToR ενώ τα μπλε βέλη υποδεικνύουν τις διαδρομές κυκλοφορίας από την διεπαφή E3 έως την διεπαφή E4 που βρίσκονται και οι δύο εντός του ToR 2.



**Εικόνα 110:** Η δομή του επιπέδου δεδομένων στο ToR 2 για το σενάριο intra-ToR

Η Εικόνα 111 απεικονίζει την εκτέλεση της εντολής *nload* η οποία «τρέχει» ταυτόχρονα σε όλους τους εμπλεκόμενους διακομιστές (δηλαδή στους E3 και E4). Με αυτόν τον τρόπο γίνεται παρακολούθηση της εισερχόμενης και της εξερχόμενης κίνησης σε κάθε διακομιστή. Πιο συγκεκριμένα, στην Εικόνα 111 (α) φαίνεται η επικοινωνία μεταξύ των διακομιστών E3 και E4. Ο διακομιστής E3 (10.2.2.1) στέλνει κίνηση δεδομένων στον διακομιστή E4 (10.2.2.129). Το εύρος ζώνης που επιτυγχάνεται σε αυτή την περίπτωση είναι σχεδόν 1,3 GBit/s. Στην Εικόνα 111 (β) η κίνηση ακολουθεί την ίδια διαδρομή, αλλά αυτή τη φορά, περισσότερες χρονοθυρίδες (slots) εκχωρούνται από τον ελεγκτή του δικτύου και αυτό έχει ως αποτέλεσμα να αυξάνεται σχεδόν κατά δύο φορές το εύρος ζώνης. Επίσης, στο συγκεκριμένο σενάριο δεν εμπλέκεται καθόλου το οπτικό τμήμα του rod switch. Η κίνηση παραμένει μόνο στο ηλεκτρικό τμήμα του ToR switch, αφού οι εμπλεκόμενοι servers ανήκουν στο ίδιο rack.



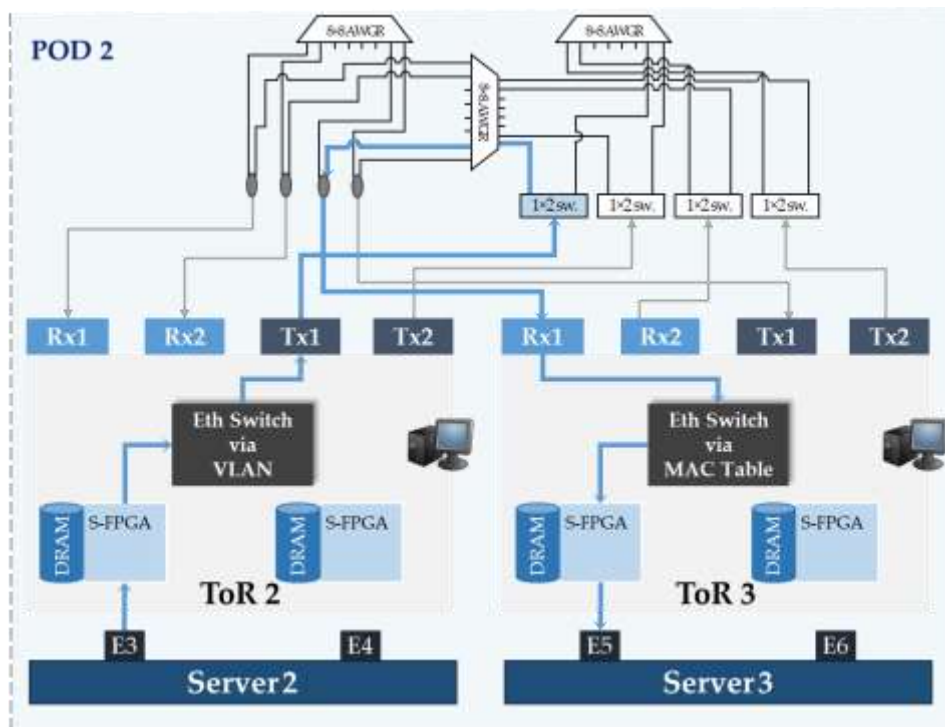
*Εικόνα 111: (α) Ο διακομιστής E3 (10.2.2.1) στέλνει κίνηση δεδομένων στον διακομιστή E4 (10.2.2.129) φτάνοντας το εύρος ζώνης 1,3 GBit/s, (β) Με την απελευθέρωση περισσότερων θυρίδων, το εύρος ζώνης διπλασιάζεται.*

#### 4.3.1.2 Σενάριο Intra-Pod

Το δεύτερο σενάριο που εφαρμόστηκε στο πρωτότυπο mini-DCN αφορούσε στην επικοινωνία εντός του pod (Intra-pod). Σε αυτήν την περίπτωση, δύο ToR που βρίσκονται εντός του pod 2, δημιουργούν μια διαδρομή επικοινωνίας μεταξύ τους. Η κίνηση αυτή ξεκινάει από τον server E3 του ToR 2 με προορισμό τον server E5 που ανήκει στον ToR 3, και τα δύο ToRs βρίσκονται εντός του pod 2. Το πείραμα πραγματοποιήθηκε σε τρία διαφορετικά μοτίβα δρομολόγησης των πακέτων ανά slot: 20, 40 και 70 slots από τα συνολικά 80 της Περιόδου Προγραμματισμού (scheduling period) του δικτύου. Με αυτόν τον τρόπο, το εύρος ζώνης αυξάνεται και αλλάζει δυναμικά. Οι εντολές για την αύξηση των καταναμημένων slots και άρα για αύξηση της κίνησης αποστέλλονται από τον ελεγκτή SDN στα FPGA που εκτελούν τις οδηγίες σε πραγματικό χρόνο ρυθμίζοντας κατάλληλα όλα τα υποσυστήματα.

Σε αυτό το πείραμα εμπλέκονται τρία μέρη του data-plane: δύο ToR switches και ένας pod switch. Στην Εικόνα 112 παρουσιάζεται ένα σχηματικό διάγραμμα της διαδρομής των πακέτων μέσω του data-plane. Ακολουθώντας τα μπλε βέλη, τα πλαίσια Ethernet που προέρχονται από τον διακομιστή E3 αποθηκεύονται πρώτα στο S-FPGA του ToR 2. Έπειτα,

η κίνηση δρομολογείται μέσω του Ethernet μεταγωγέα στο σωστό N-FPGA ως πύλη προορισμού στον οπτικό δακτύλιο του δικτύου.

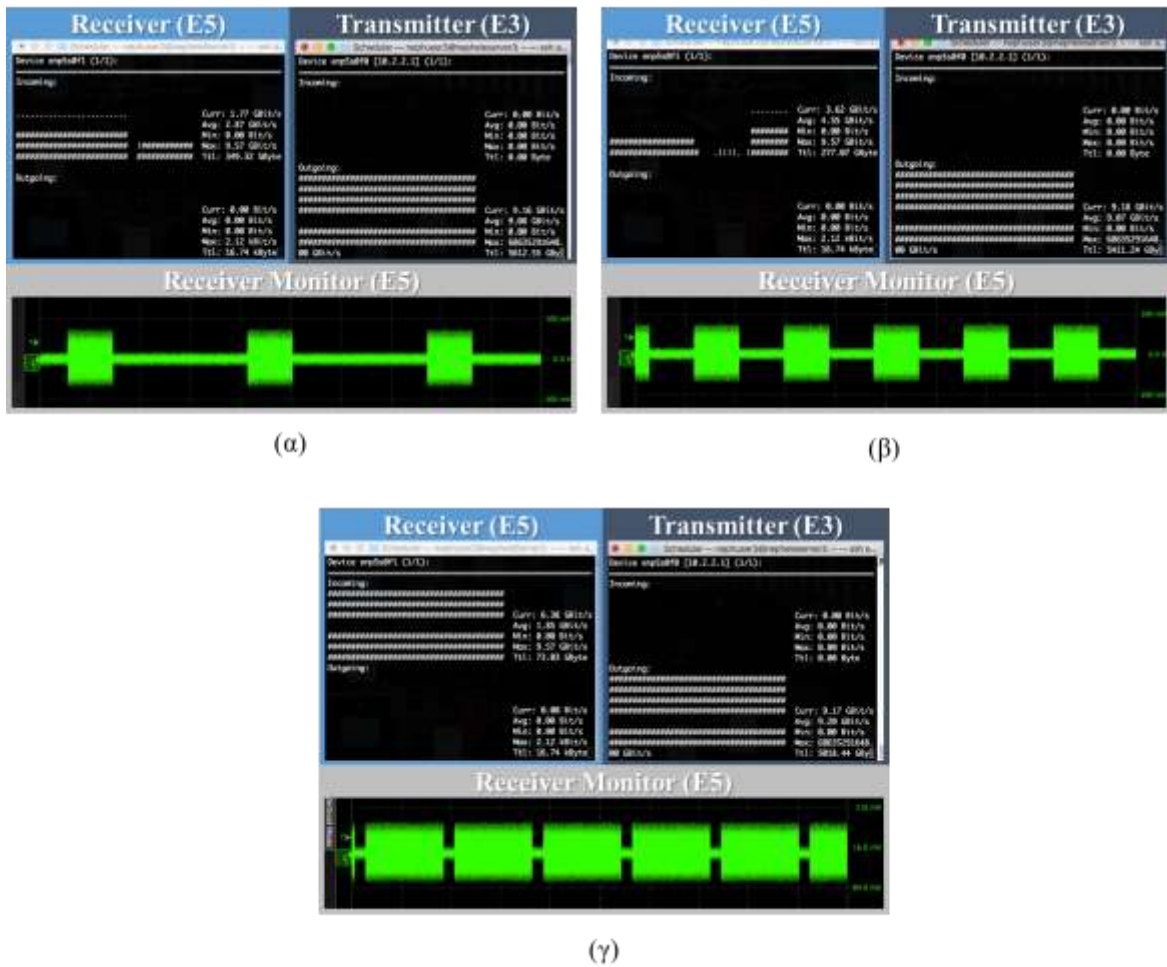


**Εικόνα 112:** Η ολοκληρωμένη διάταξη του intra-pod σεναρίου. Τα πακέτα επικοινωνίας μεταξύ ToR 2 και ToR 3 ακολουθούν τη διαδρομή που επισημαίνεται με τα μπλε βέλη

Ο προσαρμοζόμενος πομπός του ToR 2 ρυθμίστηκε έτσι ώστε να μεταδίδει την κίνηση στο κατάλληλο μήκος κύματος σύμφωνα με τον ToR προορισμό, που στην περίπτωση αυτή είναι το  $\lambda_3=1548.515\text{nm}$  (σύμφωνα με τον Πίνακα 11). Εν συνεχεία, τα πακέτα δρομολογούνται μέσω του 1x2 οπτικού διακόπτη καθορισμού, ο οποίος είναι υπεύθυνος να διαχωρίζει την intra-pod κίνηση από την inter-pod. Σε αυτήν την περίπτωση, ο οπτικός διακόπτης 1x2, δρομολογεί τα πακέτα στην intra-pod έξοδο και κατόπιν μέσω του παθητικού 8x8 κυκλικού AWG δρομολογητή, τα πακέτα προωθούνται στην ίνα που φτάνει στον δέκτη του ToR3. Είναι σημαντικό να αναφέρουμε ότι λόγω της βελτιστοποίησης του συστήματος στην οπτική διαδρομή (και σε όλα τα οπτικά μονοπάτια intra-pod) δεν υπάρχει ανάγκη για οπτική ενίσχυση.

Μετά την επιτυχή οπτική λήψη των πακέτων από τον δέκτη στο N-FPGA του ToR 3, τα εξαγόμενα Ethernet πλαίσια προωθούνται στον μεταγωγέα Ethernet, ο οποίος με τη σειρά του δρομολογεί τα πλαίσια στο S-FPGA χρησιμοποιώντας την διεύθυνση MAC του προορισμού. Τέλος, χωρίς καμία επιπλέον καθυστέρηση ή επεξεργασία, τα Ethernet πακέτα φτάνουν στον προορισμό τους που είναι ο server E5.





**Εικόνα 113:** Στιγμιότυπα των πακέτων που αποστέλλονται από τον διακομιστή E3 (10.2.2.1 ToR2) και λαμβάνονται από τον διακομιστή ToR3 E5 (10.2.3.1): (α) Σε αυτήν την περίπτωση το 25% της περιόδου προγραμματισμού (20 slots από τα 80 μιας περιόδου) καταλαμβάνεται για τη διαδρομή επικοινωνίας μεταξύ ToR2 και ToR3 επιτυγχάνοντας εύρος ζώνης 1,8 Gbit/s. (β) Σε αυτό το μισό της περιόδου (40 από τα 80 slots) καταλαμβάνεται για τη διαδρομή επικοινωνίας μεταξύ ToR2 και ToR3 επιτυγχάνοντας εύρος ζώνης 3,6 Gb/s. (γ) Σε αυτήν την περίπτωση, το 87,5% της περιόδου προγραμματισμού (70 από τα 80 slots) καταλαμβάνεται για τη διαδρομή επικοινωνίας μεταξύ ToR2 και ToR3 επιτυγχάνοντας εύρος ζώνης 6,4 Gbit/s.

Τα στιγμιότυπα της 113 δείχνουν την κίνηση που ταξιδεύει μέσω του δικτύου και έχουν ληφθεί με τη βοήθεια ενός παλμογράφου σε πραγματικό χρόνο. Πιο συγκεκριμένα καθένα στιγμιότυπο αντιπροσωπεύει διαφορετικά ποσοστά κατανομής των slots. Είναι χρήσιμο να επαναλάβουμε ότι η περίοδος προγραμματισμού του δικτύου χωρίζεται σε 80 slots (200 us κάθε slot) με χρόνο παύσης (10 us) μεταξύ τους. Προφανώς, η αύξηση των slots που κατανέμονται για μια συγκεκριμένη σύνδεση έχει ως αποτέλεσμα την άμεση αύξηση του διαθέσιμου εύρους ζώνης μεταξύ των δύο τελικών server. Στο πάνω μέρος των (α), (β) και

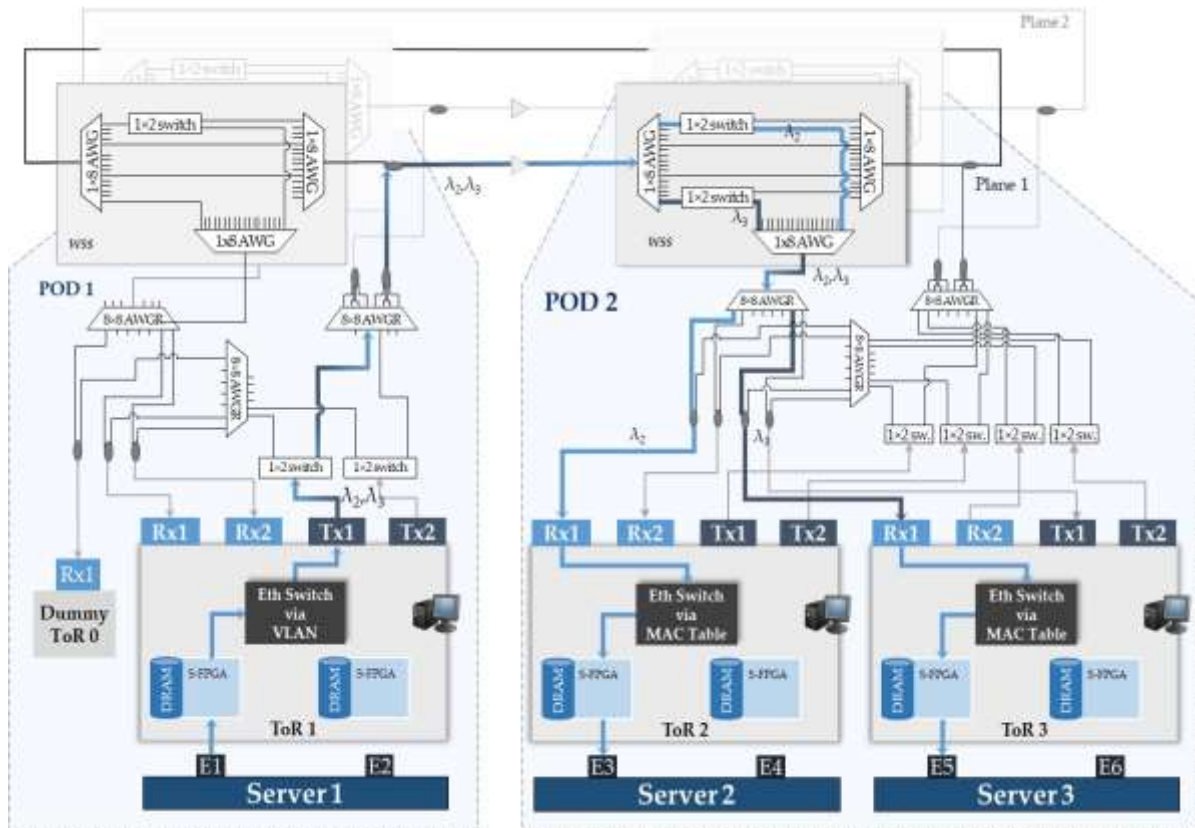
(γ) της εικόνας 113 παρουσιάζεται η εντολή *nload* που εκτελείται σε όλους τους εμπλεκόμενους servers. Το εύρος ζώνης κλιμακώνεται σύμφωνα με την κατανομή των slots. Πιο συγκεκριμένα, στην περίπτωση (α) το 25% της περιόδου προγραμματισμού (20 slots από τα 80 μιας περιόδου) καταλαμβάνεται για τη διαδρομή επικοινωνίας μεταξύ ToR2 και ToR3 επιτυγχάνοντας εύρος ζώνης 1,8 GBit/s. Στο (β) το μισό της περιόδου (40 από τα 80 slots) καταλαμβάνεται για τη διαδρομή επικοινωνίας μεταξύ ToR2 και ToR3 επιτυγχάνοντας εύρος ζώνης 3,6 Gb/s. Τέλος, στην περίπτωση (γ) το 87,5% της περιόδου προγραμματισμού (70 από τα 80 slots) καταλαμβάνεται για τη διαδρομή επικοινωνίας μεταξύ ToR2 και ToR3 επιτυγχάνοντας εύρος ζώνης 6,4 Gbit/s.

#### 4.3.1.3 Σενάριο Inter-Pod I

Η πειραματική διάταξη που χρησιμοποιήθηκε για την επικοινωνία μεταξύ των ToRs που βρίσκονται σε διαφορετικά pods φαίνεται στην Εικόνα 114. Για αυτήν την ενότητα, ακολουθούμε την διαδρομή μόνο των γαλάζιων βέλων στο σχήμα 114, αφού τα σκούρα μπλε αντιστοιχούν σε επόμενο σενάριο. Εν προκειμένω ο server E1 που βρίσκεται μέσα στο ToR 1 του pod 1 επιδιώκει επικοινωνία με τον server E3 που είναι συνδεδεμένος στον ToR 2 του pod 2. Έτσι, σε αυτό το σενάριο εμπλέκονται δύο pod switches και δύο ToR switches. Επίσης, το σενάριο κίνησης που υλοποιήθηκε περιλαμβάνει 60 από τα 80 slots της συνολικής περιόδου προγραμματισμού.

Ο προσαρμοζόμενος πομπός είναι ρυθμισμένος ώστε να εκπέμπει το μήκος κύματος που θα φτάσει το ToR 2, δηλαδή το μήκος κύματος  $\lambda_2 = 1547.715\text{nm}$ . Η μεταδιδόμενη κίνηση διέρχεται από όλα τα οπτικά στοιχεία των δύο διαδοχικών pod switches: 1×2 οπτικό διακόπτη καθορισμού (intra, inter-), 8x8 AWGR, EDFA, WSS, 8x8 AWGR και έναν συζεύκτη.

Η Εικόνα 115 απεικονίζει τα στιγμιότυπα των πακέτων που λαμβάνονται μέσω ενός παλμογράφου σε πραγματικό χρόνο και αντιπροσωπεύουν σταθερή κατανομή slots (60 από τα 80 του scheduling period). Επιπλέον, στα (α) και (β) φαίνεται η εντολή *nload* που εκτελείται στους δύο διακομιστές και δείχνει ότι το εύρος ζώνης που επιτυγχάνεται μεταξύ E1 και E3 είναι 5,12 GBit/s.



**Εικόνα 114:** Πειραματική διάταξη για το σενάριο *inter-pod*. Οι σκούρες και οι ανοιχτές μπλε γραμμές αποκαλύπτουν τις οπτικές διαδρομές από τον server E1 προς τους server E3 και E5 αντίστοιχα.



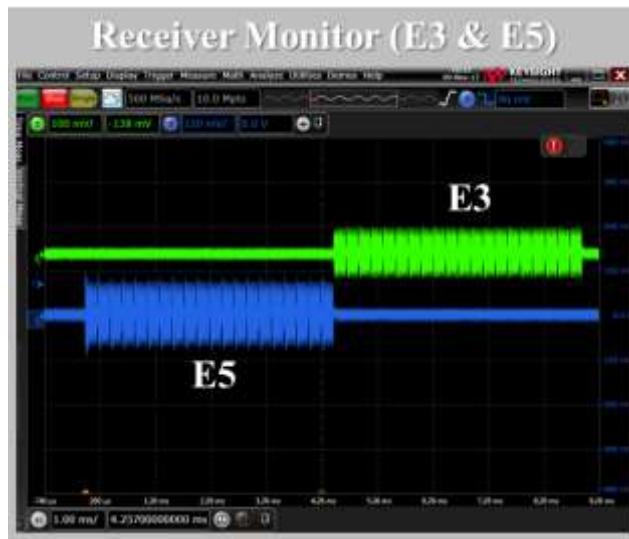
**Εικόνα 115:** (α) Τα πακέτα που αποστέλλονται από τον server E1 (10.1.1.1) του ToR1 και λαμβάνονται από το server E3 (10.2.2.1) του ToR2. Σε αυτήν την περίπτωση, το 75% της περιόδου προγραμματισμού (60 από τα 80 slots) καταλαμβάνεται για τη διαδρομή επικοινωνίας μεταξύ ToR1 και ToR2 επιτυγχάνοντας εύρος ζώνης 5,12 Gb/s.

#### 4.3.1.4 Inter-Pod scenario II

Σε αυτό το σενάριο, η επικοινωνία ακολουθεί επίσης μια inter-pod διαδρομή δηλαδή μεταξύ δύο pods. Σε αυτήν την περίπτωση όμως, ο server του ToR1, ο οποίος βρίσκεται μέσα στο pod 1, στέλνει κίνηση σε δύο διαφορετικούς ToRs στο pod 2. Πιο συγκεκριμένα, ο server E1 (ToR 1 – pod 1) δημιουργεί και στέλνει κίνηση στους servers E3 (ToR 2 – pod 2) και E5 (ToR 3 – pod 2) εναλλάξ. Ο server E1 στέλνει πλαίσια Ethernet και για τους δύο προορισμούς εναλλάξ, ενώ οι εντολές προγραμματισμού του δικτύου, που αποστέλλονται από τον ελεγκτή SDN, ορίζουν τα ανάλογα slots μέσα στην περίοδο προγραμματισμού.

Η πειραματική διάταξη για αυτό το σενάριο inter-pod με τα τρία εμπλεκόμενα ToRs απεικονίζεται και πάλι στην Εικόνα 114. Η κίνηση που προέρχεται από τον E1 μεταδίδεται οπτικά από το Tx1 του ToR 1 και προωθείται στον οπτικό δακτύλιο του επιπέδου 1 (plane 1) μέσω του 1×2 οπτικού διακόπτη καθορισμού (intra- ή inter-pod) και έναν οπτικό δρομολογητή 8×8 AWGR. Όπως αναφέρθηκε και προηγουμένως, ο προσαρμοζόμενος πομπός Tx1 δημιουργεί εναλλάξ δύο ροές δεδομένων εγγεγραμμένα σε δύο μήκη κύματος σύμφωνα με το ToR προορισμού, δηλαδή  $\lambda_2 = 1547,715 \text{ nm}$  για ToR 2 και  $\lambda_3 = 1548,515 \text{ nm}$  για ToR 3. Στην Εικόνα 114, οι δύο αυτές οπτικές διαδρομές απεικονίζονται με γαλάζιο για το ToR2 και με σκούρο μπλε για το ToR 3. Ένας οπτικός ενισχυτής (EDFA) χρησιμοποιείται για την αντιστάθμιση των οπτικών απωλειών μεταξύ των δύο pod.

Ο διακόπτης επιλογής μήκους κύματος (WSS) του pod 2, που είναι η υπεύθυνη μονάδα για την «επιλογή» συγκεκριμένων μηκών κύματος από τους δακτυλίους, προωθεί τις οπτικές ροές προς τους ToR του σύμφωνα με τις εντολές του ελεγκτή SDN. Ένα πρόσθετο παθητικό AWGR, όπως φαίνεται στην Εικόνα 114, δρομολογεί τελικά τα πακέτα στους ToR 2 και ToR 3 αντίστοιχα.



**Εικόνα 116:** Στιγμιότυπα των πακέτων που προέρχονται από το ToR1 και ελήφθησαν από στο ToR 2 (μπλε ίχνος) και στο ToR 3 (πράσινο ίχνος). Η συνολική κίνηση διαιρείται εξίσου στους δύο προορισμούς (20slots φτάνουν σε κάθε ToR).

Το πράσινο ίχνος που απεικονίζεται στην Εικόνα 116 αντιπροσωπεύει τα πακέτα που φτάνουν στο ToR 2 ενώ το μπλε τα πακέτα που λαμβάνονται από τον ToR 3. Είναι εμφανές ότι η κίνηση διαμοιράζεται σε κάθε προορισμό και πιο συγκεκριμένα το 25% της περιόδου προγραμματισμού (20 από τα 80 slots) αντιστοιχούν σε καθένα από τους δύο προορισμούς, όπως προκύπτει και από τις εικόνες 117. Τέλος, το εύρος ζώνης έφτασε σχεδόν το 1,4 Gb/s και για τους δύο servers E3 και E5.



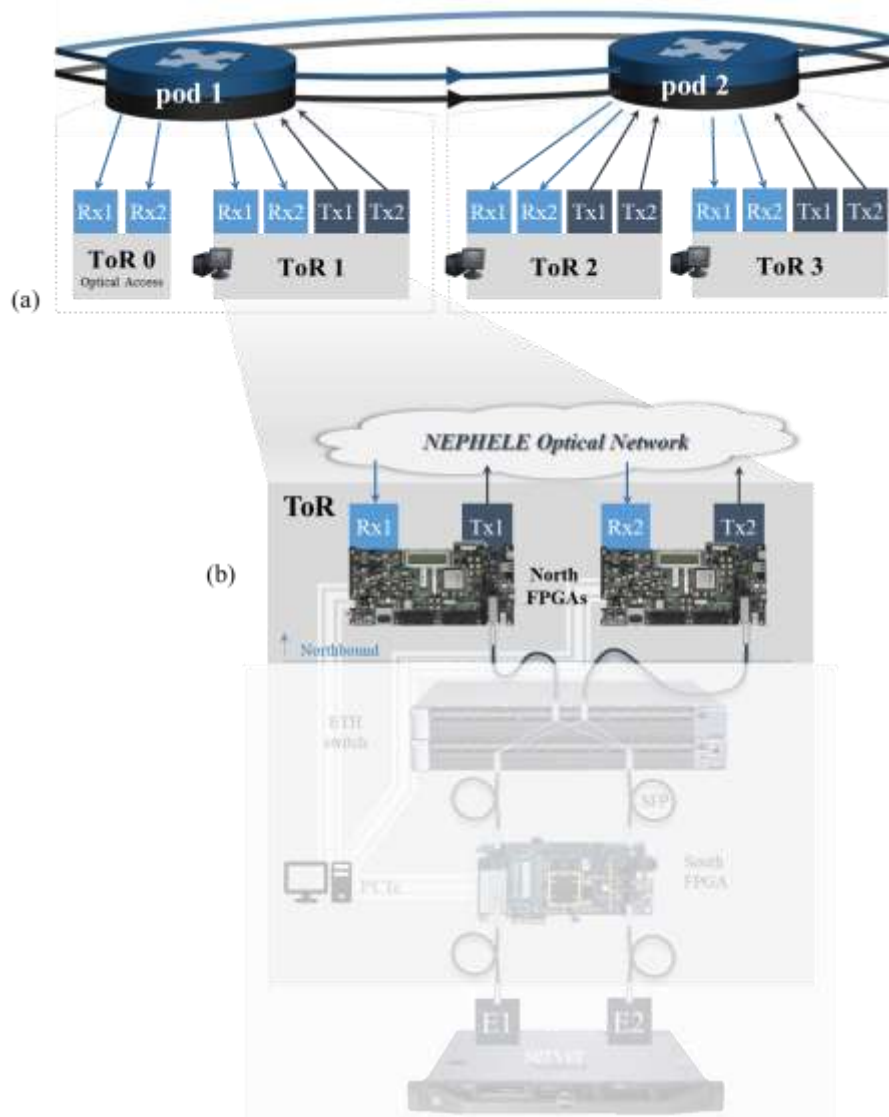
**Εικόνα 117:** (α) Ο server E1 (10.1.1.1) στέλνει κίνηση δεδομένων (β) στον server E3 (10.2.2.1) και (γ) server E5 (10.2.2.129) επιτυγχάνοντας εύρος ζώνης σχεδόν 1,4 Gb/s σε κάθε server.

#### 4.3.1.5 Συνδυασμός Intra-pod and Inter-pod σεναριών

Οι υποενότητες που ακολουθούν παρουσιάζουν επιπλέον, πιο σύνθετα σεσνάρια επικοινωνίας που εφαρμόστηκαν στο πρωτότυπο mini-DCN. Συνδυασμοί διαφόρων σεναρίων επικοινωνίας υλοποιήθηκαν προκειμένου να επιβεβαιωθεί η σταθερότητα των υπόσυστημάτων παρά την πρόσθετη πολυπλοκότητα.

Αρχικά, τα σεσνάρια που παρουσιάζονται στις επόμενες ενότητες πραγματοποιήθηκαν χρησιμοποιώντας την ίδια πειραματική διάταξη που φαίνεται για πληρότητα στην Εικόνα 118 (α) με μια κύρια διαφορά από τα προηγούμενα: η επικοινωνία μεταξύ των ToRs πραγματοποιείται από το N-FPGA του ToR switch όπως απεικονίζεται στην Εικόνα 118 (β). Αυτό σημαίνει ότι οι μονάδες που βρίσκονται κάτω από τα S-FPGA δεν εμπλέκονται σε αυτά τα πειράματα και για αυτό φαίνονται αγνά στο σχηματικό.

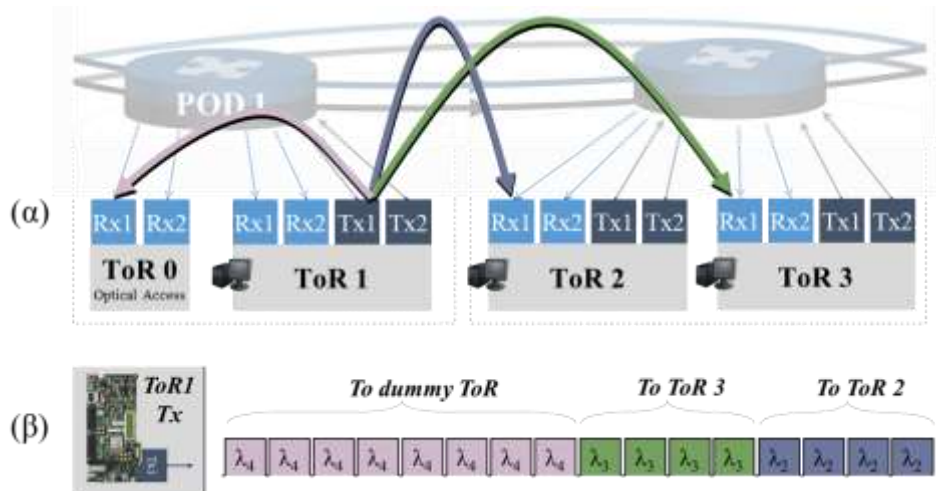
Πιο συγκεκριμένα, το N-FPGA δημιουργεί «εικονικά» πακέτα με ψευδοτυχαίες ψηφιακές ακολουθίες PRBS αντί για πραγματικά δεδομένα Ethernet. Αυτή η εικονική κίνηση πληροί όλες τις προδιαγραφές της προτεινόμενης αρχιτεκτονικής όσον αφορά το χρονοδιάγραμμα και τον ρυθμό μετάδοσης. Επίσης, το N-FPGA παράγει και παρέχει τα σήματα ελέγχου (TTL) για όλες τα οπτικά υποσυστήματα που εμπλέκονται: 1×2 οπτικοί διακόπτες καθορισμού και το 1×2 WSS που περιλαμβάνει πολλαπλούς 1×2 οπτικούς διακόπτες.



**Εικόνα 118:** (α) Η πειραματική διάταξη που χρησιμοποιείται για την αξιολόγηση των συνδυασμών *Intra-pod* και *Inter-pod* σεναρίων, (β) Το σχηματικών των εμπλεκόμενων διεπαφών του *data-plane* των *ToR* switches.

#### 4.3.1.5.1 Συνδυασμένο σενάριο 1

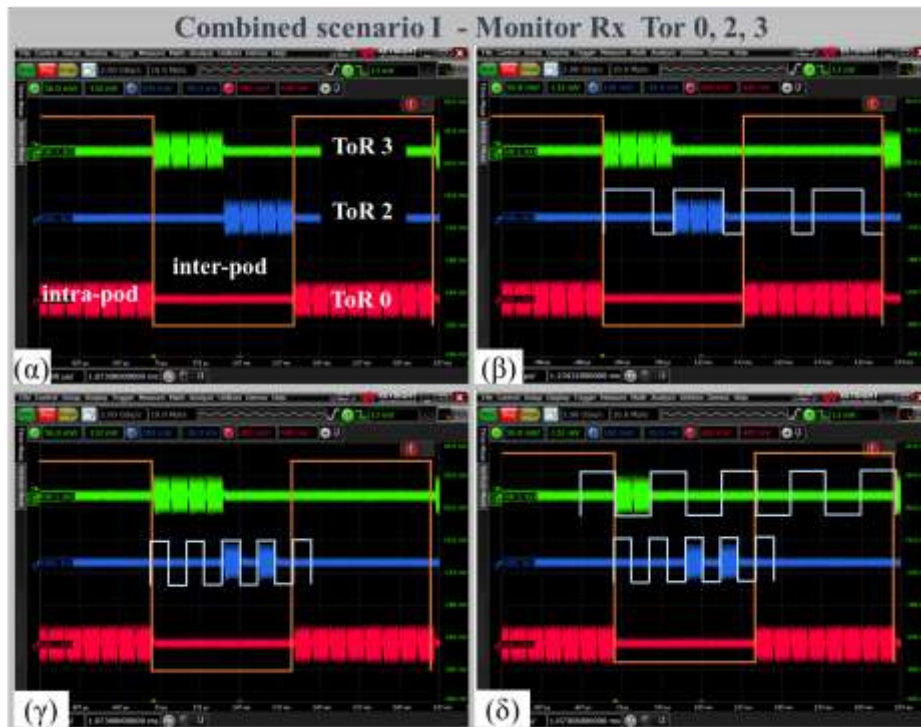
Το πρώτο σενάριο συνδυάζει την επικοινωνία εντός και μεταξύ των δύο pod, όπως φαίνεται στην Εικόνα 119 (α). Σύμφωνα με αυτό το σχηματικό, το *ToR 1* του pod 1 στέλνει πακέτα προς το *ToR 0* του pod 1 (*intra-pod*) και στα *ToR 2* και *ToR 3* του pod 2 (*inter-pod*). Ως εκ τούτου, ο προσαρμοζόμενο πομπός του *ToR 1* έχει ρυθμιστεί να μεταδίδει δεκαέξι επαναλαμβανόμενα πακέτα γραμμένα σε διαφορετικά μήκη κύματος (δηλαδή στα  $\lambda_4$ ,  $\lambda_2$  και  $\lambda_3$ ) όπως φαίνεται στην Εικόνα 119 (β).



**Εικόνα 119:** (α) Συνδεδασμένο σενάριο I. (β) Η επαναλαμβανόμενη ακολουθία πακέτων σε τρία μήκη κύματος που παράγεται από τον προσαρμοζόμενο Tx του ToR 1: οκτώ πακέτα στο  $\lambda_4$ , τέσσερα στο  $\lambda_2$  και τέσσερα στο  $\lambda_3$ .

Η Εικόνα 120 απεικονίζει τα στιγμιότυπα από τους δέκτες των ToRs και τα οποία καταγράφηκαν μέσω ενός παλμογράφου. Τα κανάλια του παλμογράφου συνδέθηκαν ταυτόχρονα με καθένα από τους τρεις φωτοδέκτες των τριών ToRs και έτσι το κόκκινο ίχνος αντιστοιχεί στο ToR 0, το μπλε στο ToR 2 και το πράσινο στο ToR 3.



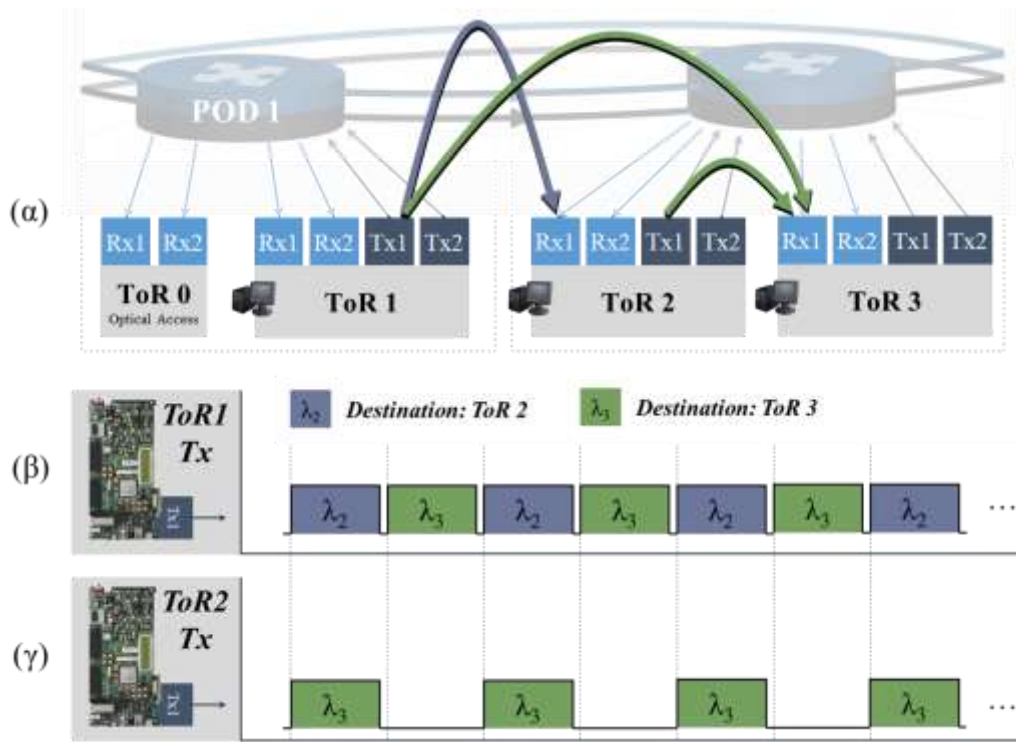


**Εικόνα 120:** Στιγμιότυπα των ληφθέντων πακέτων σε κάθε ToR. Το κόκκινο ίχνος αντιστοιχεί στον δέκτη του ToR 0, το μπλε ίχνος στον δέκτη ToR 2 και το πράσινο ίχνος στον δέκτη ToR 3.

Σε όλα τα στιγμιότυπα (α-δ) το πορτοκαλί ίχνος αντιπροσωπεύει το σήμα ελέγχου που εφαρμόζεται στον 1×2 οπτικό διακόπτη καθορισμού που διακρίνει την intra- και inter-pod επικοινωνία στο pod 1. Επιπλέον, τα λευκά ίχνη στα στιγμιότυπα οθόνης (β), (γ) και (δ) αντιπροσωπεύουν το σήμα οδήγησης για τους οπτικούς διακόπτες 1×2 που βρίσκονται μέσα στο WSS του pod 2 και αντιστοιχούν στα λ<sub>2</sub> και λ<sub>3</sub>.

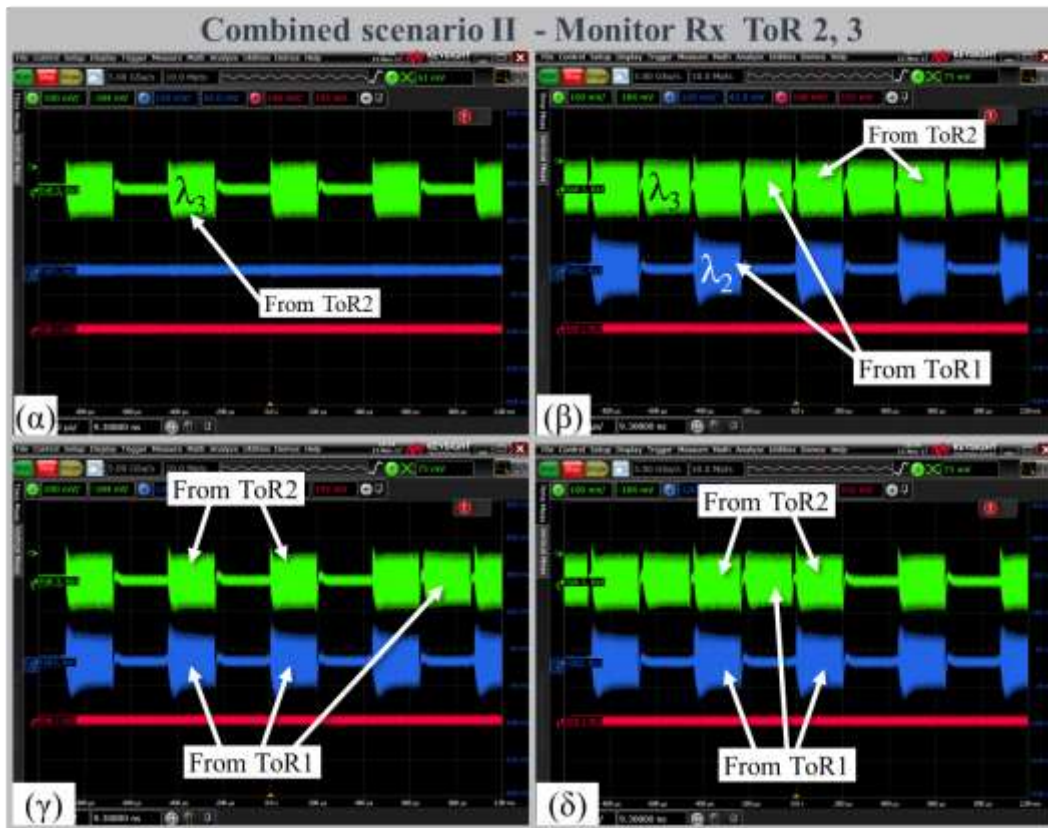
#### 4.3.1.5.2 Συνδυασμένο Σενάριο II

Το δεύτερο συνδυασμένο σενάριο που υλοποιήθηκε παρουσιάζεται στην Εικόνα 121 (α). Όπως φαίνεται, το ToR 1 του pod 1 μεταδίδει τα πακέτα στο ToR 2 και το ToR 3 του pod 2 (inter-pod). Ταυτόχρονα και συγχρονισμένα, ο πομπός του ToR 2 στέλνει κίνηση στο ToR 3 (επικοινωνία intra-pod). Ως αποτέλεσμα, ο δέκτης του ToR 3 λαμβάνει κίνηση που προέρχεται τόσο από τον ToR 1 όσο και από το ToR 2.



**Εικόνα 121:** (α) Σχηματικό για το συνδυασμένο σενάριο II Intra-pod και Inter-pod, (β) Επαναλαμβανόμενη ακολουθία πακέτων που μεταδίδονται από το ToR 1 και (γ) από το ToR 2 ταυτόχρονα και συγχρονισμένα στα  $\lambda_2$  και  $\lambda_3$ .

Η κίνηση που δημιουργείται από το ToR 1 (pod 1) και το ToR 2 (pod 2) απεικονίζεται στην Εικόνα 121 (β) και (γ) αντίστοιχα. Πιο συγκεκριμένα, όπως δείχνει η Εικόνα 121 (β) ο προσαρμοζόμενος πομπός του ToR 1 είναι ικανός να παράγει μια ακολουθία πακέτων στα  $\lambda_2$  και  $\lambda_3$ . Ομοίως, ο πομπός ToR 2 είναι προγραμματισμένος να δημιουργεί πακέτα στο  $\lambda_3$  με ένα κενό slot ενδιάμεσα, όπως φαίνεται στην Εικόνα 121 (γ), προκειμένου να αποφευχθεί οποιαδήποτε πιθανή σύγχυση με τα πακέτα στο  $\lambda_3$  που μεταδίδονται από το ToR 1.



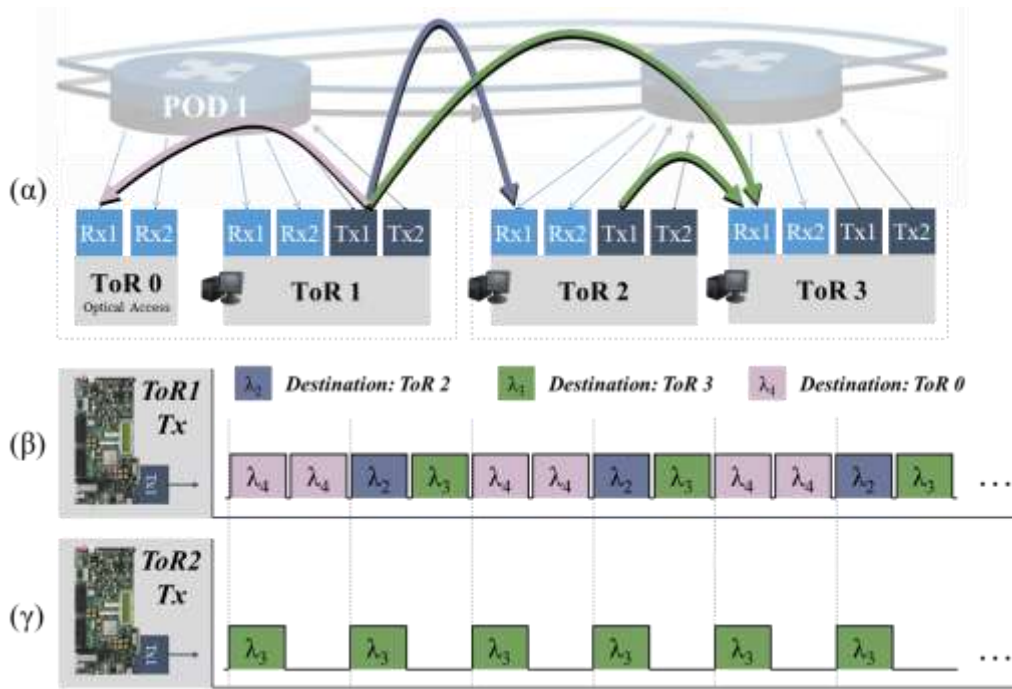
**Εικόνα 122:** Στιγμιότυπα των ληφθέντων πακέτων στους δέκτες των ToR 2 (μπλε πακέτα) και ToR 3 (πράσινα πακέτα).

Οι συνδυασμοί των προαναφερθέντων ροών πακέτων παρουσιάζονται στα στιγμιότυπα (α-δ) στην Εικόνα 122. Πιο συγκεκριμένα, η Εικόνα 122 (α) δείχνει τα πακέτα στο  $\lambda_3$  που φτάνουν στο ToR 3 και μεταδίδονται από τον ToR 2, χωρίς καμία άλλη μεταδιδόμενη κίνηση. Στη συνέχεια, ο πομπός του ToR 1 αρχίζει να στέλνει πακέτα στα μήκη κύματος  $\lambda_2$  και  $\lambda_3$  όπως φαίνεται στην Εικόνα 122 (β). Σε αυτήν την περίπτωση, ο διακόπτης  $1 \times 2$  του rod 1 προωθεί την inter-rod κίνηση και το WSS του rod 2 «κρατάει» τα συγκεκριμένα μήκη κύματος. Τα επόμενα δύο στιγμιότυπα (γ) και (δ) στην Εικόνα 122 δείχνουν περιπτώσεις στις οποίες το WSS του rod 2 «κρατάει» μερικά από τα πακέτα στο  $\lambda_3$ .

#### 4.3.1.5.3 Συνδυασμένο Σενάριο III

Το τρίτο και τελευταίο σενάριο που πραγματοποιήθηκε στο πρωτότυπο mini-DCN είναι ένας συνδυασμός των δύο προηγούμενων σεναρίων και φαίνεται στο σχηματικό σχήμα της Εικόνας 123 (α). Σε αυτό το σενάριο η κίνηση που παράγεται από το ToR1 είναι τόσο intra-rod (ToR 0 - rod 1) όσο και inter-rod (ToR 2 και ToR 3 – rod 2). Εν τω μεταξύ, το ToR 2 δημιουργεί μια άλλη intra-rod ροή, αλλά αυτή τη φορά μέσα στο rod 2 μεταδίδοντας

πακέτα με προορισμό το ToR 3. Επιπλέον, σε αυτό το σενάριο εμπλέκονται τρία μήκη κύματος ( $\lambda_2$ ,  $\lambda_3$  και  $\lambda_4$ ) και προφανώς οι δύο πομποί είναι συγχρονισμένοι.



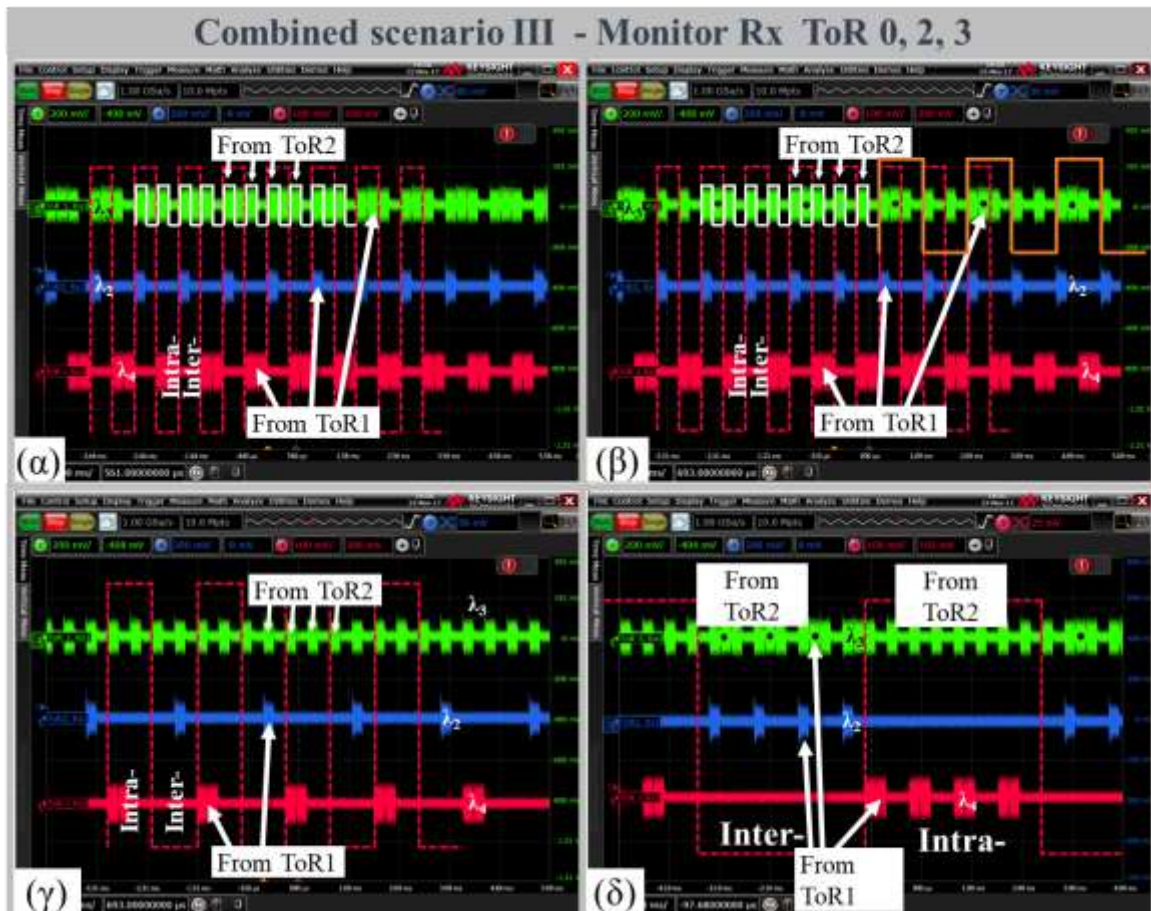
**Εικόνα 123:** (α) Σχηματικό για το συνδυασμένο σενάριο III, (β) Επαναλαμβανόμενη ακολουθία πακέτων που μεταδίδονται από το ToR 1 και (γ) από το ToR 2 ταυτόχρονα και συγχρονισμένα στα  $\lambda_2$ ,  $\lambda_3$  και  $\lambda_4$ .

Η κίνηση που δημιουργείται από τους δύο προσαρμοζόμενους πομπούς απεικονίζεται στην Εικόνα 123 (β) και (γ). Ο πομπός του ToR 1 (pod 1) δημιουργεί μια ακολουθία πακέτων που είναι γραμμένα στα μήκη κύματος  $\lambda_4$ ,  $\lambda_2$  και  $\lambda_3$ . Όπως συνέβη στο προηγούμενο συνδυασμένο σενάριο, ο πομπός του ToR 2 (pod 2) προγραμματίζεται ώστε να παράγει πακέτα στο  $\lambda_3$  με ένα κενό slot ανάμεσά τους.

Η Εικόνα 124 δείχνει τέσσερα στιγμιότυπα των πακέτων που φτάνουν στους τρεις διαφορετικούς ToR. Τα κόκκινα πακέτα αντιπροσωπεύουν αυτά που είναι γραμμένα στο μήκος κύματος  $\lambda_4$  και που λαμβάνονται από τον δέκτη του ToR 0. Ομοίως, τα μπλε και πράσινα ίχνη δείχνουν τα πακέτα που φτάνουν στους δέκτες των ToR 2 και ToR 3 αντίστοιχα. Είναι σημαντικό να αναφέρουμε ότι σε όλα τα στιγμιότυπα οι δύο πομποί έχουν ρυθμιστεί να εκπέμπουν τις ακολουθίες που φαίνονται στην Εικόνα 123 (β) και (γ) ταυτόχρονα και συγχρονισμένα.

Επιπλέον, σε όλα τα στιγμιότυπα (α-δ) στην Εικόνα 124, η διακεκομμένη κόκκινη κυματομορφή υποδηλώνει τον διαχωρισμό intra- και inter-pod επικοινωνίας στο pod 1. Στα ίδια στιγμιότυπα τα λευκά βέλη με τις αντίστοιχες ετικέτες υποδεικνύουν το ToR

προέλευσης κάθε πακέτου. Πιο συγκεκριμένα, στην Εικόνα 124 (α) και (β) τα πακέτα που μεταδίδονται από το ToR 2 (intra-rod εντός του rod 2) φαίνονται με πράσινο χρώμα. Επίσης, το πορτοκαλί ίχνος και οι μαύρες κουκκίδες που επισημαίνονται Εικόνα 124 (β) και (δ) αντιπροσωπεύουν την κατάσταση ενεργοποίησης/απενεργοποίησης του WSS για το  $\lambda_3$  και τα πακέτα που μεταδίδονται από το ToR 1 αντίστοιχα.



**Εικόνα 124:** Στιγμιότυπα των ληφθέντων πακέτων στους δέκτες των ToR 0, 2 και 3. Το κόκκινο ίχνος αντιστοιχεί στον δέκτη ToR 0 (dummy ToR), το μπλε ίχνος στον δέκτη ToR 2 και το πράσινο ίχνος στον δέκτη ToR 3.

#### 4.3.2 Επικοινωνία Inter-Datacenter μεταξύ Αθήνας (Ελλάδα) και Πίζας (Ιταλία)

Προκειμένου να επαληθεύσουμε την ικανότητα της προτεινόμενης αρχιτεκτονικής δικτύου να ελέγχει ετερογενείς τεχνολογίες, έγινε επαλήθευση μέσω δοκιμών που πραγματοποιήθηκαν μεταξύ ενός εικονικού mini-datacenter το οποίο αναπτύχθηκε στην Πίζα της Ιταλίας και του πρωτοτύπου mini-DCN στην Αθήνα.

Αυτό το σύστημα περιλαμβάνει δύο δίκτυα datacenter που προσομοιώνονται μέσω Mininet, συνδεδεμένα μέσω ενός δικτύου DC (επίσης εξομοίωσης) σε ένα τρίτο φυσικό DCN. Τα

δύο πρώτα δίκτυα εκτελούνται σε Virtual machines (VM) σε εγκαταστάσεις στην Πίζα, ενώ το φυσικό DCN είναι το πρωτότυπο mini-DCN που περιγράφεται σε αυτό το κεφάλαιο. Οι δύο τοποθεσίες διασυνδέονται μέσω VPN που δημιουργείται μεταξύ δύο κεντρικών υπολογιστών που τα διαχειρίζονται, όπως φαίνεται στην Εικόνα 125 (α).



**Εικόνα 125:** (α) Το δίκτυο επικοινωνίας για intra-DC και inter-DC δοκιμές (β) Το εύρος ζώνης που επιτεύχθηκε μέσω του διαδικτύου και μετρήθηκε στα 4,91 Mbit/s.

Το πείραμα εκτελείται μέσω του NIDO GUI που στέλνει ένα αίτημα για μια end-to-end διαδρομή μεταξύ ενός εξομοιωμένου server που βρίσκεται στην Πίζα και ενός φυσικού server στο πρωτότυπο mini-DCN που βρίσκεται στο ΕΦΕ την Αθήνα.

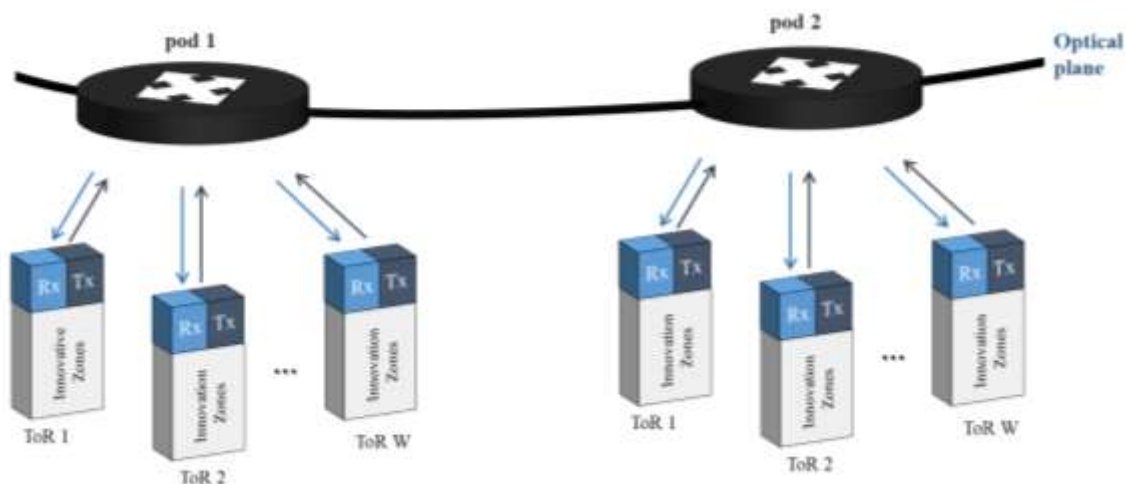
Για να επαληθευθεί η πραγματική συνδεσιμότητα μεταξύ των δύο server, χρησιμοποιήθηκε το εργαλείο *iPerf* για να δημιουργηθεί πραγματική κίνηση μεταξύ των servers, ενώ μέσω του *tcpdump* επαληθεύθηκε η κίνηση αυτή. Αν και το εύρος ζώνης που μετρήθηκε (4,91 Mbit/s - Εικόνα 125 (β)) είναι πολύ χαμηλό για τα πρότυπα της προτεινόμενης αρχιτεκτονικής, η κίνηση μεταφέρεται μέσω του VPN και μέσω του διαδικτύου, επομένως ο πραγματικός ρυθμός μεταφοράς δεν αποτελεί σημαντική ένδειξη της απόδοσης του φυσικού πρωτότυπου DCN.

# ΚΕΦΑΛΑΙΟ 5

## Σύνοψη των αποτελεσμάτων και προτάσεις για μελλοντική έρευνα

### 5.1 Σύνοψη των αποτελεσμάτων

Η παρούσα διδακτορική διατριβή είχε σαν κύριο αντικείμενό της την μελέτη μιας υβριδικής αρχιτεκτονικής δικτύου για Κέντρα Δεδομένων και των υποσυστημάτων της, που συνδυάζει και αξιοποιεί τα πλεονεκτήματα των κλασικών ηλεκτρικών διασυνδέσεων με αυτά των οπτικών. Στο πλαίσιο αυτό η πειραματική αξιολόγηση έγινε μέσω διαφόρων τεχνικών οπτικής δρομολόγησης. Σε αντίθεση με τις κλασικές fat-tree τοπολογίες που εφαρμόζονται στα περισσότερα δίκτυα των κέντρων δεδομένων, η προτεινόμενη υβριδική αρχιτεκτονική επιτρέπει την οριζόντια επικοινωνία μεταξύ των servers χωρίς την γραμμική αύξηση των ενδιάμεσων επιπέδων και κόμβων.



*Εικόνα 126: Σχηματικό της προτεινόμενης υβριδικής αρχιτεκτονικής*

Συνοψίζοντας, όπως φαίνεται στην Εικόνα 126 το κύριο δομικό στοιχείο της αρχιτεκτονικής είναι το pod, που φιλοξενεί έναν αριθμό από racks, που με τη σειρά τους εξυπηρετούν μερικές χιλιάδες από servers. Τα pods διασυνδέονται μεταξύ τους μέσω ενός οπτικού δακτυλίου. Κάθε rack είναι υπό τη διαχείριση ενός top-of-rack μεταγωγέα (ToR switch), ενώ όλοι οι ToR switch διασυνδέονται με το pod switch με οπτικό τρόπο σε μια

τοπολογία αστέρα. Κάθε ToR switch είναι εξοπλισμένο με ένα προσαρμοζόμενο οπτικό πομπό και έναν οπτικό δέκτη που υποστηρίζει την λειτουργία πακέτων. Για την κίνηση που προορίζεται εντός του ίδιου pod (intra-pod), η οπτική δρομολόγηση πραγματοποιείται κυρίως μέσω παθητικών στοιχείων οπτικού φιλτραρίσματος. Η επικοινωνία μεταξύ διακομιστών που βρίσκονται μεταξύ διαφορετικών pods (inter-pod) είναι ένας συνδυασμός μεταγωγής μήκους κύματος (wavelength switching) αλλά και χώρου (space switching), επιτρέποντας την επαναχρησιμοποίηση των μηκών κύματος μεταξύ των pods.

Η είσοδος ή η έξοδος της κίνησης προς και από τον οπτικό δακτύλιο βασίζεται στο μήκος κύματος και η τεχνική δρομολόγησης ονομάζεται add/drop multiplexing. Το επίπεδο δεδομένων (dataplane) της αρχιτεκτονικής λειτουργεί με βάση την τεχνική TDMA με χρονοσχισμές ή αλλιώς timeslots, όπου τα «slots» είναι τα χρονικά τμήματα στα οποία είναι δυνατή η επικοινωνία μεταξύ των racks (rack-to-rack). Τα slots (και επομένως οι πόροι δικτύου) μπορούν να διατεθούν δυναμικά σε επικοινωνία μεταξύ των rack και έτσι το δίκτυο μπορεί να πετύχει σχεδόν πλήρη χρήση της χωρητικότητάς του, οδηγώντας σε μείωση της κατανάλωσης της ενέργειας αλλά και κόστους.

Η μεγαλύτερη πρόκληση στα πλαίσια της πολύμηνης μελέτης της προτεινόμενης αρχιτεκτονικής ήταν αφενός η υλοποίηση των επιμέρους οπτικών υποσυστημάτων και αφετέρου η συγχρονισμένη ενοποίησή τους. Έτσι τα σημαντικότερα αποτελέσματα που προέκυψαν από τον σχεδιασμό, υλοποίηση και πειραματική αξιολόγηση των της αρχιτεκτονικής παρουσιάζονται σε αυτήν την ενότητα.

### **5.1.1 Υλοποίηση και πειραματική αξιολόγηση του γρήγορου προσαρμοζόμενου πομπού του ToR switch.**

Ο δυναμικά προσαρμοζόμενος πομπός του ToR switch, όπως έχει αναφερθεί, χρησιμεύει για την αποτύπωση των ψηφιακών δεδομένων των endhosts σε έναν οπτικό φορέα. Το μήκος κύματος φέρει τις πληροφορίες δρομολόγησης για τη συγκεκριμένη ροή και επιλέγεται από ένα προκαθορισμένη μπάντα μηκών κύματος γύρω από τη C-band. Τα δύο δομικά στοιχεία του πομπού είναι η οπτική πηγή (laser) και ο οπτικός διαμορφωτής (Modulator), καθώς και τα κυκλώματα ελέγχου για την επιθυμητή ρύθμιση του laser.

Έγινε πλήρης χαρακτηρισμός του συντονισμού του μήκους κύματος του λέιζερ (τεχνολογίας Modulated Grating Y-branch structure, MG-Y) στη φασματική περιοχή της C μπάντας, και πιο συγκεκριμένα στην περιοχή μεταξύ 1529,55nm - 1561,1nm. Σχετικά με την οπτική ισχύ που εκπέμπει το λέιζερ, οι μετρήσεις έδειξαν ότι σε κάθε συχνότητα η οπτική ισχύς ήταν τουλάχιστον 8 dBm, ενώ παρατηρήθηκε SMSR > 40 dB για κάθε μέτρηση. Σχετικά με τον χρόνο που απαιτείται ώστε το λέιζερ να ρυθμιστεί από το ένα μήκος κύματος σε ένα άλλο, η πειραματική διαδικασία ανέδειξε ότι αυτό επιτυγχάνεται στην χειρότερη περίπτωση μέσα σε ~8 ns, επιβεβαιώνοντας αφενός την επιτυχημένη



προσαρμογή του λέιζερ στην ηλεκτρονική πλακέτα οδήγησης (PCB) κι αφετέρου την πολύ γρήγορη απόκρισή του στην αλλαγή κατάστασης.

Περαιτέρω και προκειμένου να γίνει αξιολόγηση της ποιότητας του οπτικού σήματος εξόδου του λέιζερ, καταγράφηκαν μετρήσεις διαμορφωμένου σήματος για κάθε ένα από τα δύο εναλλασσόμενα μήκη κύματος ξεχωριστά. Μέσω των καμπύλων του ρυθμού BER έναντι της λαμβανόμενης οπτικής ισχύος για δύο μήκη κύματος, που έχουν φασματική απόσταση > 18,5 nm, φαίνεται ότι έχουμε επιτυχή λειτουργία χωρίς σφάλματα και όμοια συμπεριφορά στα ίδια λαμβανόμενα επίπεδα ισχύος, επιβεβαιώνοντας ότι το λέιζερ είναι ικανό να εναλλάσσει το μήκος κύματος του σε ταχύτητα σημαντικά υψηλή, διατηρώντας την ακεραιότητα του σήματος.

Για να επιτευχθεί ο δυναμικός συντονισμός του λέιζερ χρησιμοποιήθηκε ένας μετατροπέας ψηφιακού σήματος σε αναλογικό (Digital-to-Analog Converter DAC), όπως περιγράφεται στην ενότητα 3.1 παρέχοντας με δυναμικό τρόπο τα ρεύματα στους τομείς της φάσης και των ανακλαστήρων του λέιζερ. Ο DAC ενσωματώνεται πάνω στο FPGA μέσω ενός FMC adapter προκειμένου να λάβει τα απαραίτητα ψηφιακά σήματα που αντιστοιχούν στα αναλογικά ρεύματα. Είναι σημαντικό να αναφερθεί ότι για τον υπολογισμό του συνολικού χρόνου μεταγωγής του πομπού στο ToR switch, πρέπει να ληφθεί υπόψη, εκτός από την ταχύτητα εναλλαγής μήκους κύματος του λέιζερ (που όπως είπαμε ήταν ~8ns) και η συμπεριφορά του DAC, δηλαδή ο χρόνος ο οποίος απαιτείται για τη μετάβαση από μια στάθμη ρεύματος σε μία άλλη. Η μετάβαση αυτή μετρήθηκε να είναι σχεδόν 10 ns και συμφωνεί με τις προδιαγραφές του κατασκευαστή του DAC.

Έπειτα διερευνήθηκε ο συνολικός χρόνος μεταγωγής και συντονισμού του πομπού από ένα οποιοδήποτε κανάλι σε ένα άλλο. Ο Πίνακας 3, που επαναλαμβάνεται και παρακάτω για ευκολία, παρουσιάζει όλους τους σχετικούς χρόνους μετάβασης με τη χρήση του DAC, οι οποίοι μετρήθηκαν να είναι από 9.4ns έως 20.1ns.

Τύπος μετάβασης	Κανάλια	Χρόνος μετάβασης μαζί με τον χρόνο μετάβασης του DAC	Αναφορά [46] (χωρίς τον χρόνο μετάβασης του DAC)
Μέσα στον ίδιο ρυθμό	51 → 36	11.2 ns	3 ns
Εντός του ίδιου υπερ-ρυθμού σε διαδοχικούς ρυθμούς	51 → 52	14.2 ns	4 ns
Εντός του ίδιου υπερ-ρυθμού σε μη διαδοχικούς ρυθμούς	51 → 61	18.2 ns	7 ns
Σε διαδοχικούς υπερ-ρυθμούς	51 → 22	9.4 ns	0.4 ns
Σε μη διαδοχικούς υπερ-ρυθμούς	51 → 13	20.1 ns	12 ns

### 5.1.2 Χαρακτηρισμός και πειραματική αξιολόγηση των γρήγορων οπτικών μεταγωγέων

Ο 4×4 PLZT διακόπτης που μελετήθηκε αποτελείται από τέσσερα στάδια μεταγωγής που επιτυγχάνονται μέσω Mach-Zehnder συμβολόμετρων. Μέσω αυτών των σταδίων μπορούν να επιτευχθούν είκοσι-τέσσερις πιθανές καταστάσεις. Έτσι, εφαρμόζοντας TTL σήματα ελέγχου σε κάθε MZ, μπορεί να υλοποιηθεί η επιθυμητή οπτική διαδρομή. Πραγματοποιήθηκαν τρία διαφορετικά πειράματα προκειμένου να αξιολογηθεί στατικά η απόδοση του γρήγορου 4×4 PLZT. Ο στόχος του πρώτου πειράματος ήταν να διερευνηθούν οι απώλειες στην οπτική ισχύ που εισάγονται όταν το φως διέρχεται μέσα από τον οπτικό διακόπτη. Τα αποτελέσματα του πειράματος έδειξαν ότι οι απώλειες του μεταγωγέα κυμαίνονται από 13,5 dB έως 18,1 dB, με μέση τιμή τα 14,6 dB. Αξίζει επίσης να σημειωθεί ότι φαίνεται ότι υπάρχει σημαντική εξάρτηση των απωλειών ανάλογα με την πόλωση του εισερχόμενου οπτικού φορέα η οποία κατά μέσο όρο μετρήθηκε στα 2,5dB. Σε σύγκριση με τις τυπικές προδιαγραφές του διακόπτη, οι μετρήσεις αποκλίνουν σημαντικά τόσο για τις απώλειες όσο και για την εξάρτησή τους από την πόλωση του εισερχόμενου φωτός.

Το δεύτερο πείραμα αφορούσε στη μέτρηση του χρόνου μεταγωγής του 4×4 PLZT, δηλαδή στον χρόνο που μεσολαβεί μεταξύ δύο διαφορετικών σταθερών καταστάσεων του διακόπτη. Για να μετρηθεί ο χρόνος μεταγωγής χρησιμοποιήθηκε μια φωτοδίοδος συνδεδεμένη με έναν παλμογράφο πραγματικού χρόνου. Με βάση τους χρόνους ανόδου και πτώσης της οπτικής ισχύος στις θύρες εξόδου μετρήθηκε πειραματικά ο μέσος χρόνος μεταγωγής και ήταν περίπου 33 ns. Επιπλέον, κατά τη διάρκεια του πειράματος έγιναν μετρήσεις του χρόνου μεταγωγής μεταβάλλοντας την οπτική πόλωση στο σήμα εισόδου και το συμπέρασμα είναι ότι δεν παρατηρήθηκε καμία αξιοσημείωτη εξάρτηση του χρόνου μεταγωγής από την πόλωση του εισερχόμενου οπτικού σήματος. Αν και η μετρηθείσα τιμή των 33ns είναι σημαντικά πιο αργή από την αναμενόμενη τιμή σύμφωνα με τις τυπικές προδιαγραφές (τυπικά 10 ns), το αποτέλεσμα είναι εντός των αποδεκτών τιμών της υπό διερεύνηση αρχιτεκτονικής της διατριβής.

Το τρίτο πείραμα του στατικού χαρακτηρισμού του 4×4 PLZT οπτικού διακόπτη αφορούσε στη μέτρηση της παρεμβολής (crosstalk) μεταξύ των καναλιών εξόδου. Για το λόγο αυτό, ο διακόπτης ρυθμίστηκε στατικά σε διάφορες οπτικές διαδρομές και η οπτική ισχύς ανιχνεύθηκε και καταγράφηκε μέσω ενός αναλυτή οπτικού φάσματος. Ο λόγος παρεμβολής μετρήθηκε στα 26 dB, ελαφρώς χειρότερος από την τυπική τιμή (30dB).

Το επόμενο στάδιο της πειραματικής αξιολόγησης πραγματοποιήθηκε με σκοπό την αξιολόγηση του οπτικού τμήματος του ToR switch με τη χρήση του 4×4 PLZT οπτικού

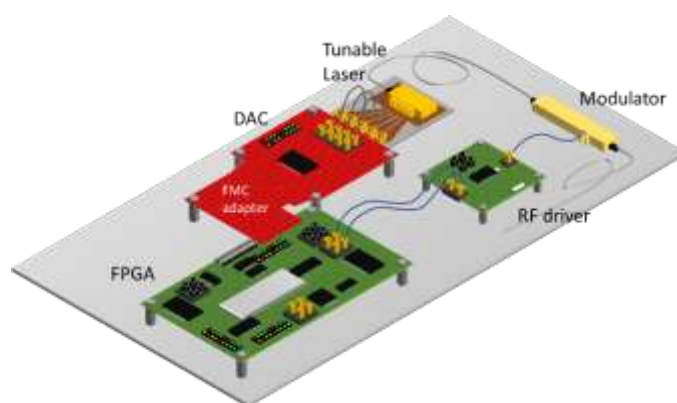
διακόπτη. Ο διακόπτης 4×4 τοποθετείται στην έξοδο κάθε πομπού και πριν από κάθε δέκτη και στην πραγματικότητα εξυπηρετεί δύο κύριες λειτουργίες: από την πλευρά του πομπού δρομολογεί την εξερχόμενη κίνηση από τους ToR είτε ως *intra-ToR* είτε *inter-ToR* ενώ από την άλλη πλευρά διαχειρίζεται την εισερχόμενη κυκλοφορία που προέρχεται από το pod switch και τη δρομολογεί προς τα κατάλληλα τερματικά. Κάθε πομπός στέλνει σε διαφορετικά επίπεδα του δικτύου ανάλογα με το μήκος κύματος που θα εκπέμψει. Επιπλέον, το ToR switch είναι εξοπλισμένο και με ίσο αριθμό φωτοδεκτών, έτσι ώστε να μπορεί να λαμβάνει σήματα από τα διαφορετικά επίπεδα δικτύου.

Σχετικά με την *intra-ToR* επικοινωνία, για λαμβανόμενη ισχύ μεγαλύτερη από - 6,67 dBm δεν παρατηρήθηκαν σφάλματα (error-free transmission), που αντιστοιχούν σε BER καλύτερο από  $3 \times 10^{-13}$ . Επιπλέον για *intra-pod* επικοινωνία και για λαμβανόμενη ισχύ μεγαλύτερη από -6,64 dBm δεν παρατηρήθηκαν σφάλματα, που αντιστοιχούν σε BER καλύτερο από  $3 \cdot 10^{-13}$ , ενώ για *inter-pod* επικοινωνία δεν παρατηρήθηκαν σφάλματα για οπτική ισχύ μεγαλύτερη από -4,82 dBm (BER καλύτερο από  $3 \cdot 10^{-13}$ ).

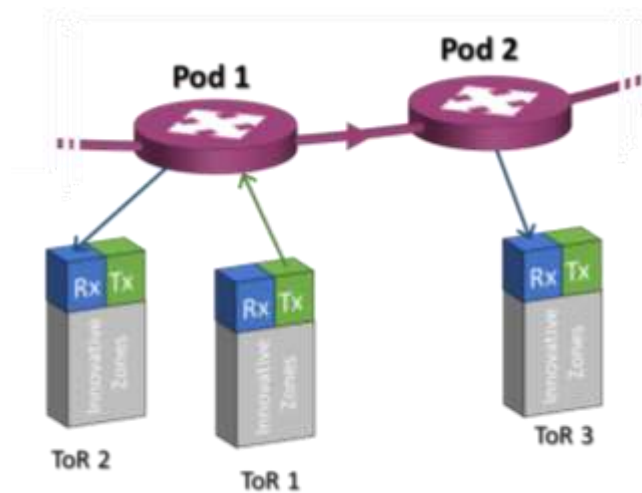
### 5.1.3 Ενοποίηση και πειραματική αξιολόγηση οπτικών υποσυστημάτων του πρωτότυπου μίνι-DataCenter στο Εργαστήριο Φωτονικών Επικοινωνιών (ΕΦΕ)

Σε επόμενο στάδιο έγινε η ενοποίηση των οπτικών υποσυστημάτων της προτεινόμενης αρχιτεκτονικής, που παρουσιάζεται στο ΚΕΦΑΛΑΙΟ 3, με ορίζοντα την δημιουργία και επιτυχή λειτουργία του πρωτότυπου μίνι-Datacenter στο Εργαστήριο Φωτονικών Επικοινωνιών (ΕΦΕ) που περιγράφεται στο ΚΕΦΑΛΑΙΟ 4.

Πρώτα έγινε η ενοποίηση του Προσαρμοζόμενου Πομπού του ToR switch και αξιολογήθηκε ο δυναμικός έλεγχός του μέσω FPGA. Έτσι, λοιπόν δείχτηκε η «συναρμολόγηση» του πομπού από τα επιμέρους στοιχεία που το απαρτίζουν, δηλαδή το ρυθμιζόμενο λέιζερ μαζί με τον ψηφιακό μετατροπέα (DAC), τον ενισχυτή σήματος RF και του οπτικού διαμορφωτή. Έγινε πλήρης χαρακτηρισμός του υποσυστήματος και ακολούθησε η αξιολόγηση της δυναμικής του λειτουργίας μέσω FPGA. Σε αυτό το πλαίσιο καθορίστηκαν τα μήκη κύματος που θα εκπέμψει ενώ παράλληλα έγινε πρόβλεψη για την σύγχρονη και αρμονική λειτουργία του με τα υπόλοιπα στοιχεία της αρχιτεκτονικής.

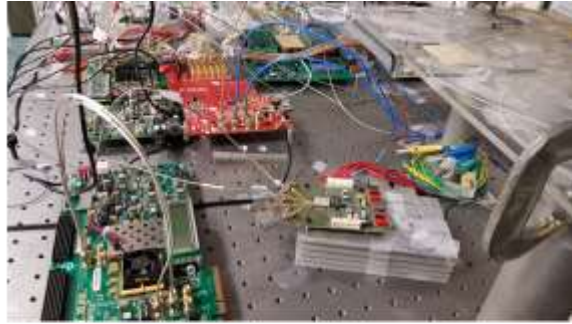


Ακολούθησαν αρκετά πειράματα και εφαρμόστηκαν διαφορετικά σενάρια επικοινωνίας σε ένα οπτικό δίκτυο στο οποίο έχουν συνδυαστεί ToR και pod switches μαζί με όλα τα παθητικά και ενεργά οπτικά στοιχεία, που προσομοιάζει στο δίκτυο της προτεινόμενης αρχιτεκτονικής. Τα πειράματα διεξήχθησαν στη βάση του απλοποιημένου γενικού σχηματικού που επαναλαμβάνεται για πληρότητα στην Εικόνα 127 και περιλάμβαναν δύο pod switches (pod 1 και pod 2) και τρεις ToR switches (ToR 1, ToR 2 και ToR 3), έτσι ώστε να μπορούν να πραγματοποιηθούν και να δοκιμαστούν τόσο τα σενάρια επικοινωνίας εντός όσο και μεταξύ των pod, καθώς και οι συνδυασμοί τους. Σε όλα τα σενάρια δρομολόγησης τα επίπεδα της λαμβανόμενης ισχύος στους δέκτες ήταν χωρίς σφάλματα παρόλη την πολυπλοκότητα της οπτικής διάταξης. Τα πειραματικά αποτελέσματα επιδεικνύουν αφενός την επιτυχημένη ενσωμάτωση και διασύνδεση όλων των οπτικών υποσυστημάτων του ToR και του pod switch και αφετέρου την επιτυχημένη λειτουργία των τεχνικών οπτικής δρομολόγησης.



**Εικόνα 127:** Επισκόπηση του απλοποιημένου σχηματικού για τις πειραματικές δοκιμές

Το στάδιο της ενοποίησης όλων των υποσυστημάτων της προτεινόμενης αρχιτεκτονικής ήταν η «συναρμολόγηση» του πρωτότυπου μίνι DataCenter στο ΕΦΕ. Στιγμές από την πολύμηνη διαδικασία αποτυπώνονται στις φωτογραφίες που παρουσιάζονται σε αυτήν την ενότητα.

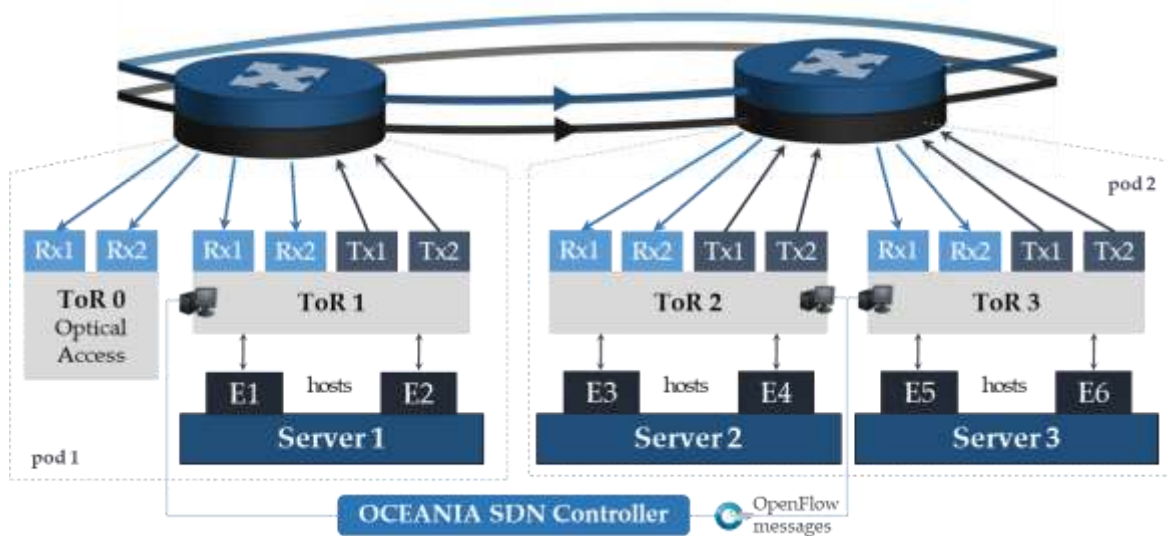


Η πολυπλοκότητα της δομής απαιτούσε εντατική προσπάθεια τεχνικής δουλειάς αλλά και πολύωρων συζητήσεων για την ερμηνεία και μελέτη των δυσκολιών που παρουσιάζονταν

σε κάθε βήμα. Παρόλα αυτά η αρμονική συνεργασία όλων μας στο ΕΦΕ είχε σαν αποτέλεσμα την πετυχημένη κατασκευή σε μια ενιαία οπτική δομή του πρωτότυπου.



Σκοπός του πρωτότυπου mini-DCN, όπως έχει ήδη αναφερθεί, ήταν να αναδειχθεί ο αρμονικός συντονισμός όλων των επιμέρους δομικών μερών του από το φυσικό στρώμα (physical layer) μέχρι το επίπεδο δικτύου (network orchestration), ενώ χρησιμοποιήθηκε για να γίνει απόδειξη του ότι η προτεινόμενη αρχιτεκτονική μπορεί αν δουλέψει αποτελεσματικά.



**Εικόνα 128:** Το διάγραμμα της αρχιτεκτονικής δομής του πρωτότυπου datacenter

Με τη δομή που παρουσιάζεται στην Εικόνα 128 προσομοιώνονται δύο οπτικά επίπεδα (optical planes) με δύο pod στα πρότυπα της προτεινόμενης αρχιτεκτονικής.

Όσον αφορά στο φυσικό επίπεδο, η επικοινωνία μεταξύ των διακομιστών στο προτεινόμενο δίκτυο επιτυγχάνεται με τη χρήση διαφορετικών μηκών κύματος για τη δρομολόγηση των πακέτων, αξιοποιώντας το συνδυασμένο αποτέλεσμα των γρήγορων λέιζερ και των παθητικών οπτικών δρομολογητών AWGR. Από την άλλη πλευρά, όσον αφορά το στρώμα δικτύου/πρωτοκόλλου, σε κάθε διεπαφή Ethernet εκχωρούνται διαφορετικές IP διευθύνσεις.

Όσον αφορά στο επίπεδο ελέγχου, ο ελεγκτής SDN και ο ενορχηστρωτής του πρωτότυπου μίνι datacenter εκτελούνται σε ένα απομακρυσμένο υπολογιστή που είναι συνδεδεμένος μέσω τοπικού δικτύου. Ο ελεγκτής SDN συνδέεται με τους ανάλογους SDN agents, οι οποίοι εκτελούνται στους υπολογιστές που ελέγχουν όλα τα FPGA μέσω PCIexpress. Κατά την επίδειξη του πρωτοτύπου, οι οδηγίες που προέρχονται από το επίπεδο ελέγχου διαμορφώνονται, μεταφέρονται και εφαρμόζονται σε πραγματικό χρόνο από τα FPGA.

Πραγματοποιήθηκαν αρκετά σενάρια επικοινωνίας κατά τη λειτουργία του πρωτότυπου μίνι datacenter από άκρο σε άκρο (end-to-end) και σε πραγματικό χρόνο (real-time). Πιο συγκεκριμένα, τα σενάρια επικοινωνίας μέσα στο πρωτότυπο mini-DCN στην υποενότητα 4.3.1, ενώ μια διασύνδεση του πρωτοτύπου της Αθήνας με ένα άλλο datacenter στην Πίζα της Ιταλίας μέσω διαδικτύου παρουσιάζεται στην υποενότητα 4.3.2.

## 5.2 Προτάσεις για μελλοντική έρευνα

Στα πλαίσια της παρούσας διατριβής παρουσιάστηκε μια ενιαία αρχιτεκτονική για οπτικά δίκτυα Κέντρων Δεδομένων μαζί με τεχνικές δρομολόγησης με τη χρήση παθητικών και ενεργητικών φωτονικών στοιχείων.

Είναι δεδομένο ότι οι επικοινωνίες μέσω οπτικών ινών είναι και θα εξακολουθούν να είναι κινητήρια δύναμη για τη φωτονική. Επομένως, η ενσωμάτωση φωτονικών στοιχείων σε Κέντρα Δεδομένων είναι απαραίτητη και πλέον αναπόσπαστη. Αρκετές ευέλικτες πλατφόρμες φωτονικών τσιπ είναι στην αγορά, ενώ τα ενσωματωμένα λέιζερ και οπτικοί μεταγωγείς χαμηλής κατανάλωσης ενεργοποιούν βασικές λειτουργίες στα datacom, με την αυξανόμενη χρήση πολλαπλών λωρίδων και του WDM.

Η περαιτέρω εξέλιξη της φωτονικής αλλά και των αρχιτεκτονικών δικτύου μπορούν να φέρουν πραγματική επανάσταση στον τρόπο λειτουργίας των Κέντρων Δεδομένων και να προσθέσουν νέες λειτουργίες και να θωρακίσουν την αξιοπιστία και την σταθερότητά τους. Στη συνέχεια παρουσιάζονται ορισμένες προτάσεις προς αυτήν την κατεύθυνση.

### 5.2.1 Οπτική Μεταγωγή στα Κέντρα Δεδομένων και εναλλακτικές αρχιτεκτονικές δικτύων

Όπως ήδη έχει αναφερθεί και στην εισαγωγή, η οπτική μεταγωγή μπορεί να μειώσει το κόστος του εύρους ζώνης στα μελλοντικά δίκτυα κέντρων δεδομένων μειώνοντας την ανάγκη για ηλεκτρονική μεταγωγή και τις αντίστοιχες οπτικές-ηλεκτρικές-οπτικές μετατροπές. Ωστόσο, οι προκλήσεις τόσο στο φυσικό επίπεδο όσο και στο επίπεδο ελέγχου παραμένουν ισχυρές και εμποδίζουν την υιοθέτηση της οπτικής μεταγωγής.

Οι οπτικοί μεταγωγείς, προκειμένου να είναι πρακτικοί στα κέντρα δεδομένων, πρέπει να πληρούν τα θεμελιώδη κριτήρια απόδοσης στο φυσικό επίπεδο σχετικά με τον αριθμό των θυρών, την ταχύτητα επαναδιαμόρφωσης και την ακεραιότητα του οπτικού σήματος. Οι οπτικοί πομποδέκτες στα σημερινά κέντρα δεδομένων δεν μπορούν να ανεχθούν μεγάλη υποβάθμιση του σήματος, επομένως είναι κρίσιμο να ελαχιστοποιηθούν οι απώλειες που εισάγονται από τους οπτικούς διακόπτες τις παρεμβολές. Οι περισσότερες διασυνδέσεις στα κέντρα δεδομένων έχουν χαμηλό περιθώριο και αντοχή στις επιπρόσθετες απώλειες (insertion losses), περιθώριο το οποίο μικραίνει όσο αυξάνονται οι ρυθμοί των δεδομένων. Οι τοπολογίες δικτύου single-optical-hop (αυτές που απαιτούν οπτικά σήματα να περνούν από έναν μόνο οπτικό διακόπτη μεταξύ ηλεκτρονικών τελικών σημείων) ελαχιστοποιούν την υποβάθμιση του σήματος, αλλά τα σύγχρονα δίκτυα κέντρων δεδομένων παρέχουν



συνδεσιμότητα μεταξύ χιλιάδων rack servers, απαιτώντας έναν αντίστοιχο αριθμό θυρών στους οπτικούς μεταγωγείς ώστε να διασυνδεθούν όλα αυτά τα rack σε μια τέτοια τοπολογία.

Οι οπτικοί διακόπτες πρέπει να επαναρυθμιστούν αρκετά γρήγορα για να διατηρήσουν την απόδοση της εφαρμογής, τουλάχιστον για τις εφαρμογές υψηλού εύρους ζώνης. Η μεταγωγή με ρυθμό πακέτων (<100 ns) θα ήταν ιδανική, αλλά οι οπτικοί διακόπτες δεν μπορούν εύκολα να αντιγράψουν την ικανότητα των ηλεκτρονικών μεταγωγέων πακέτων να αποθηκεύουν πακέτα, και έτσι συνήθως η διαχείρισή τους γίνεται από ένα εξωτερικό, κεντρικό επίπεδο ελέγχου. Στα κέντρα δεδομένων οι ταχύτητες αναδιαμόρφωσης οπτικού κυκλώματος που απαιτούνται θα πρέπει να είναι το πολύ 10-100 μs. Ωστόσο, ο προγραμματισμός οπτικών κυκλωμάτων με βέλτιστο εύρος ζώνης σε πραγματικό χρόνο είναι πολύπλοκος και πιθανότατα μη πρακτικός σε κλίμακα κέντρου δεδομένων.

Επειδή οι παραδοσιακές προσεγγίσεις σχεδιασμού των οπτικών διακοπών που βασίζονται σε beam steering και κυματοδηγούς δεν έχουν εκπληρώσει αυτούς τους στόχους απόδοσης, εξετάζονται διαφορετικές προσεγγίσεις τόσο για το σχεδιασμό οπτικών διακοπών όσο και για την αρχιτεκτονική οπτικού δικτύου. Προς αυτήν την κατεύθυνση, έχουν προταθεί μερικώς διαμορφώσιμοι οπτικοί διακόπτες, οι οποίοι, λόγω της περιορισμένης παραμετροποίησής τους, μπορούν να σχεδιαστούν για να πληρούν τον αριθμό των θυρών του κέντρου δεδομένων, την ταχύτητα αναρρύθμισης και τους στόχους απόδοσης μετάδοσης σήματος. Κατ' αντιστοιχία προτείνονται και αρχιτεκτονικές δικτύου που να βασίζονται σε αυτούς τους διακόπτες, η οποία, παρά την περιορισμένη δυνατότητα διαμόρφωσης του διακόπτη, παρέχει υψηλή απόδοση ενώ εξαλείφει την ανάγκη για έναν κεντρικό ελεγκτή για τον προγραμματισμό των κυκλωμάτων.[64]

### **5.2.2 Μελέτη γρήγορων οπτικών μεταγωγέων και περισσότερων I/Os.**

Η επικοινωνία με οπτικές ίνες εξακολουθεί να είναι η κύρια βιομηχανική κινητήρια δύναμη για τη φωτονική. Επομένως, είναι χρήσιμοι να δοκιμαστούν σε αυτή την αγορά ευέλικτες πλατφόρμες φωτονικών τσιπ. Τα ενσωματωμένα λέιζερ και οι μεταγωγείς MEMS χαμηλής κατανάλωσης μπορούν να υποστηρίξουν βασικές λειτουργίες των datacom, με την αυξανόμενη χρήση της πολυπλεξίας μήκους κύματος και της χρήσης πολλαπλών παράλληλων δομών. Παράλληλα, στις περιπτώσεις των Κέντρων Δεδομένων υπάρχει όλο και αυξανόμενη ανάγκη για μεταγωγή κυκλώματος.

Με την ανάπτυξη του Cloud Computing και της τεχνητής νοημοσύνης, τα κέντρα δεδομένων λειτουργούν ολοένα και περισσότερο ως ένα γιγαντιαίο υπολογιστικό σύστημα κι έτσι η απόδοσή του επηρεάζεται και συναρτάται από την καθυστέρηση (latency) της ηλεκτρονικής μεταγωγής πακέτων. Αυτό θα μπορούσε να βελτιωθεί σημαντικά με έναν

γρήγορο οπτικό διακόπτη/μεταγωγέα ικανό να μπορεί πολύ γρήγορα να επαναδιαμορφώσει τις διασυνδέσεις στο κέντρο δεδομένων και έτσι να υποβοηθήσει μια συγκεκριμένη εφαρμογή (π.χ. ένα πρόγραμμα εκπαίδευσης AI). Εξάλλου οι μεταγωγείς κυκλώματος είναι διαφανείς ως προς τη ροή των δεδομένων με χαμηλό κόστος ανά θύρα που υποστηρίζει υπάρχοντα πρότυπα όπως Ethernet ή Infiniband.

Σήμερα οι περισσότεροι μεταγωγείς φωτονικών κυκλωμάτων που βασίζονται σε τσιπ είτε είναι αρκετά αργοί είτε εισάγουν υψηλές απώλειες στις οπτικές διασυνδέσεις. Η χρήση MEMS μεταγωγέων και οπτικών ενισχυτών μπορεί να μετριάσει αυτές τις ανησυχίες. Ωστόσο, μελετώνται διάφορες πιθανές λύσεις με αυξανόμενη πολυπλοκότητα. Στην απλούστερη από αυτές, προτείνεται ένας διακόπτης MEMS με στάδια ενίσχυσης για να επιτευχθεί χαμηλή απώλεια ίνας προς ίνα (<3 dB) και χρόνο μεταγωγής 100μs. Σε μια δεύτερη προσέγγιση που ξεπερνά τον σχετικά αργό χρόνο απόκρισης των MEMS, εφαρμόζονται δύο παράλληλοι διακόπτες (χρησιμοποιώντας SOA) με δυνατότητα γρήγορης μεταγωγής (100ps) όπου όταν ο ένας λειτουργεί ο άλλος επαναδιαμορφώνεται. Τέλος μελετάται η πλήρης ικανότητα των πομποδεκτών να συνδυάσουν οπτικές και ηλεκτρικές θύρες για τη διασύνδεση του διακόπτη κυκλώματος με έναν ηλεκτρονικό διακόπτη πακέτων.

# Βιβλιογραφία

- [1] Cisco, "Cisco, "Cisco Annual Internet Report (2018–2023)", white paper," [Online]. Available: <https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/white-paper-c11-741490.html> . [Accessed 2021].
- [2] A. Singh and e. al., "Jupiter Rising: A Decade of Clos Topologies and Centralized Control in Google's Datacenter Network," in *SIGCOMM*, 2015.
- [3] W. M. Mellette, "A Practical Approach to Optical Switching in Data Centers," in *2019 Optical Fiber Communications Conference and Exhibition (OFC)*, San Diego, CA, USA, 2019.
- [4] D. Zang, M. Chen, N. Sun, R. Proietti, S. J. B. Yoo and Z. Cao, "OpticV: An energy-efficient datacenter network architecture by MEMS-based all-optical bypassing," in *2016 IEEE Optical Interconnects Conference (OI)*, San Diego, CA, 2016.
- [5] Y. B. Ezra and B. I. Lembrikov, "Application of All-Optical Memory for Advanced Modulation Formats in Communication Intra-Datacenter Networks (Intra-DCNs)," in *2018 20th International Conference on Transparent Optical Networks (ICTON)*, Bucharest, 2018.
- [6] Z. Zhong, C. Guo and G. Shen, "Scheduling Traffic Switching in An All-Optical Intra-Datacenter Network with Sub-Waveband Switching," in *2018 Asia Communications and Photonics Conference (ACP)*, Hangzhou, 2018.
- [7] K. Sato and e. al., "A Large-Scale Wavelength Routing Optical Switch for Data Center Networks," *IEEE Commun. Mag.*, vol. 51, no. 9, p. 46–52, Sept. 2013.

- [8] G. M. Saridis and e. al., "Lightness: A Function-Virtualizable Software Defined Data Center Network With All-Optical Circuit/Packet Switching," *Journal of Lightwave Technology*, vol. 34, no. 7, pp. 1618-1627, 2016.
- [9] X. Xue and e. al., "SDN-Enabled Reconfigurable Optical Data Center Networks Based on Nanoseconds WDM Photonics Integrated Switches," in *2019 21st International Conference on Transparent Optical Networks (ICTON)*, Angers, France, 2019.
- [10] S. Peng, B. Guo, C. Jackson, R. Nejabati, F. Agraz, S. Spadaro, G. Bernini, N. Ciulli and D. Simeonidou, "Multi-Tenant Software-Defined Hybrid Optical Switched Data Centre," *J. Lightwave Technol.*, vol. 33, no. 15, pp. 3224-3233, 1 Aug.1, 2015.
- [11] M. I. &. Strategy, "Intel's Disaggregated Server Rack," 2013.
- [12] R. Pries, M. Jarschel, D. Schlosser, M. Klopff and P. Tran-Gia, "Power Consumption Analysis of Data Center Architectures," *Green Communications and Networking. GreeNets 2011. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, Springer*, vol. 51, 2012.
- [13] T. Benson, A. Akella and D. A. Maltz, "Network traffic characteristics of datacenters in the wild," in *Conference on Internet measurement (IMC)*, pp. 267-280, 2010.
- [14] 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.
- [15] "Open Compute Project," 2014. [Online]. Available: <http://www.opencompute.org/>.
- [16] "Facebook Shatters the Computer Server Into Tiny Pieces," 2013. [Online]. Available: <http://www.wired.com/2013/01/facebook-server-pieces/>.
- [17] N. Heath, "Open Compute: Does the data center have an open future?," 2013. [Online]. Available: <http://www.zdnet.com/open-compute-does-the-data-center-have-an-open-future-7000013012/>.
- [18] N. Farrington, G. Porter, S. Radhakishnan, H. H. Bazzaz, V. Subramanya, Y. Fainman, G. Papen and A. Vahdat, "Helios: A Hybrid Electrical/Optical Switch Architecture for Modular Data Centers," in *SIGCOMM'10*, New Delhi, India, 2010.

- [19] G. Wang, D. G. Andersen, M. Kaminsky, K. Papagiannaki, T. S. E. N, M. Kozuch and M. Ryan, "c-Through: Part-time Optics in Data Centers," in *SIGCOMM'10*, New Delhi, India,, 2010.
- [20] K. Chen, A. Singla, A. Sing, K. Ramachandran, Y. Z. L. Xu, X. Wen and Y. Chen, "OSA: An Optical Switching Architecture for Data Center Networks With Unprecedented Flexibility," 2013.
- [21] C. 3. MEMS, "CALIEN : "The Software Defined Hybrid Packet"," CALIENT Technologies, 2013. [Online]. Available: <https://www.ofcconference.org/getattachment/3cb12ab7-949e-4459-afd5-a6b8a9b87d82/The-Software-Defined-Hybrid-Packet-Optical-Datacen.aspx>.
- [22] H. Liu, R. K. F. Lu, A. Forencich, G. M. Voelker, G. Papen, A. C. Snoeren and G. Porter, "REACToR: A REconfigurable pAcket and Circuit ToR Switch," in *IEEE Photonics Society Summer Topical Meetings*, USA, 2013.
- [23] K. Christodoulopoulos, K. Katrinis, M. Ruffini and D. O'Mahony, "Accelerating HPC Workloads with Dynamic Adaptation of a Software-Defined Hybrid Electronic/Optical Interconnect," in *OFC 2014*, USA, 2014.
- [24] D. Siracusa, G. Maier, V. Linzalata and A. Pattavina, "Scalability of Optical Interconnections based on the Array Waveguide Grating in High Capacity Routers," in *ONDM 2011*, Bologna, Italy, 2011.
- [25] X. Ye, Y. Yin, S. Yoo, P. Mejia, R. Proietti and V. Akella, "DOS - A Scalable Optical Switch for Datacenters," in *ACM/IEEE Symposium on Architectures for Networking and Communications Systems (ANCS)*, La Jolla, CA USA, 2010.
- [26] R. Proietti, Y. Yin, R. Yu, C. J. Nitta, V. Akella, C. Mineo and S. J. B. Yoo, "Scalable Optical Interconnect Architecture Using AWGR-Based TONAK LION Switch With Limited Number of Wavelengths," *Journal of Lightwave Technology*, vol. 31, no. 24, pp. pp. 4087-4096, 2013.
- [27] K. Xi, Y. Kao, M. Yang and H. J. Chao, "Petabit Optical Switch for Data Center Networks," 2010.
- [28] J. Gripp, J. E. Simsarian, J. D. LeGrange, P. Bernasconi and D. T. Neilson, "Photonic Terabit Routers: The IRIS Project," in *OFC/NFOEC*, San Diego, CA USA, 2010.

- [29] Y.-K. Yeo, Z. Xu, C.-Y. Liaw, D. Wang, Y. Wang and T.-H. Cheng, "A 448x448 Optical Cross-Connect for High Performance Computers and Multi-Terabit/s Routers," in *OFC/NFOEC*, San Diego, CA USA, 2010.
- [30] T. Niwa, H. Hasegawa and K.-i. Sato, "A 270 x 270 Optical Cross-connect Switch Utilizing Wavelength Routing with Cascaded AWGs," in *OFC/NFOEC*, Anaheim CA USA, 2013.
- [31] N. Farrington and e. al., "A Multiport Microsecond Optical Circuit Switch for Data Center Networking," *IEEE Photonics Technology Letters*, vol. 25, no. 16, pp. pp. 1589-1592, 2013.
- [32] N. Farrington, P.-C. Sun, A. Forencich, J. Ford, Y. Fainman, G. Papen, A. Vahdat and G. Porter, "A Demonstration of Ultra-Low-Latency Data Center Optical Circuit Switching," in *SIGCOMM'12*, Helsinki, Finland, 2012.
- [33] A. Predieri and e. al., "Lightness Deliverable D2.2: Design Document for the proposed network architecture," 2013.
- [34] R. R. Grzybowski and e. al., "The OSMOSIS Optical Packet Switch for Supercomputers: Enabling Technologies and Measured Performance," *Photonics in Switching*, pp. pp. 21-22, 2007.
- [35] P. Bakopoulos, K. Tokas, C. Spatharakis and H. Avramopoulos, "Slotted optical datacenter network with sub-wavelength resource allocation," in *IEEE Photonics Society Summer Topical Meeting Series (SUM)*, San Juan, 2017.
- [36] K. Tokas, I. Patronas, C. Spatharakis, D. Reisis, P. Bakopoulos and H. Avramopoulos, "Slotted TDMA and optically switched network for disaggregated datacenters,," in *19th International Conference on Transparent Optical Networks (ICTON)*, Girona, 2017.
- [37] G. Landi and e. al., "SDN control framework with dynamic resource assignment for slotted optical datacenter networks," in *Optical Fiber Communications Conference and Exhibition (OFC)*, Los Angeles, CA, 2017.
- [38] K. Christodoulopoulos, K. Kontodimas, A. Siokis, K. Yiannopoulos and E. Varvarigos, "Efficient bandwidth allocation in the NEPHELE optical/electrical

datacenter interconnect," *IEEE/OSA Journal of Optical Communications and Networking*, vol. 9, no. 12, pp. 1145-1160, 2017.

- [39] "NIDO orchestrator," [Online]. Available: <https://github.com/nextworks-it/nephele-nido>.
- [40] G. Landi, M. Capitani, A. Kretsis, P. Kokkinos, K. Christodoulopoulos and E. Varvarigos, "Joint Intra- and Inter-Datacenter Network Optimization and Orchestration," in *Optical Fiber Communications Conference and Exposition (OFC) 2018*, San Diego, CA, 2018.
- [41] "OCEANIA SDN controller," [Online]. Available: <https://github.com/nextworks-it/oceania-dcn-controller>.
- [42] A. Kretsis, L. Corazza, K. Christodoulopoulos, P. Kokkinos and E. Varvarigos, "An emulation environment for SDN enabled flexible IP/optical networks," in *2016 18th International Conference on Transparent Optical Networks (ICTON)*, Trento, 2016.
- [43] O. P. A. -. L. OVERVIEW. [Online]. Available: <https://www.opendaylight.org/what-we-do/current-release/lithium>.
- [44] "OpenFlow Protocol," [Online]. Available: <https://opennetworking.org/sdn-resources/customer-case-studies/openflow/>.
- [45] I. Patronas, N. Gkatzios, V. Kitsakis, D. Reisis, K. Christodoulopoulos and E. Varvarigos, "Scheduler Accelerator for TDMA Data Centers," in *26th Euromicro International Conference on Parallel, Distributed and Network-based Processing (PDP)*, Cambridge, 2018.
- [46] Finisar, "Application Note AN-2095 controlling the S7500 CW Tunable Laser," Finisar, [Online]. Available: [https://www.sevensix.co.jp/wordpress/wp-content/uploads/2016/11/an-2095-controlling\\_the\\_s7500\\_cw\\_tunable\\_laser-reva\\_updated.pdf](https://www.sevensix.co.jp/wordpress/wp-content/uploads/2016/11/an-2095-controlling_the_s7500_cw_tunable_laser-reva_updated.pdf).
- [47] G. Losio, "Laser Sources For Optical Transceivers," Prolabs White Paper:, 2004. [Online]. Available: <http://www.completenetworks.co.uk/uploads/laser-sources-optical-transceivers.pdf>.

- [48] J. Buus, M.-C. Amann and D. J. Blumenthal, "Tunable Laser Diodes and Related Optical Sources," Second Edition SPIE Press & Wiley-IEEE press, 2005, chapter 7 «Widely Tunable Monolithic Laser Diodes », [Online]. Available: <https://spie.org/samples/PM144.pdf>.
- [49] S. R. Systems, "Calibrate Steinhart-Hart Coefficients for Thermistors," Stanford Research Systems, Rev B. Revised Oct 2012. [Online]. Available: <https://www.thinksrs.com/downloads/pdfs/applicationnotes/LDC%20Note%204%20NTC%20Calculator.pdf>.
- [50] ITU-T, "Spectral grids for WDM applications: DWDM," [Online]. Available: [https://www.itu.int/rec/dologin\\_pub.asp?lang=e&id=T-REC-G.Sup42-201810-I!!PDF-E&type=items](https://www.itu.int/rec/dologin_pub.asp?lang=e&id=T-REC-G.Sup42-201810-I!!PDF-E&type=items).
- [51] J. Buus, M.-C. Amann and D. J. Blumenthal, "Widely Tunable Monolithic Laser Diodes," in *Tunable Laser Diodes and Related Optical Sources, Second Edition*, SPIE PRESS BOOK, 2005.
- [52] H. Heininger and U. Hannemann, "FPGA based Experimentation System for Widely Tunable MG-Y Branch Semiconductor Laser," *Electrorevue*, vol. 3, no. ISSN 1213-1539, 2012.
- [53] R. Todt, T. Jacke, R. Laroy, G. Morthier and M.-C. Amann, "Demonstration of Vernier effect tuning in tunable twin-guide laser diodes," *IEE Proc.-Optoelectronics*, vol. 152, no. 2, April 2005.
- [54] J. Fabrega, B. Schrenk, J. L. FB Bo, M. Forzati, P. Rigole and J. Prat, "Modulated Grating Y-Structure Tunable Laser for-Routed Networks and Optical Access," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 17, pp. 1542-1551, Nov.-Dec. 2011.
- [55] "DAC34H84 Quad-Channel, 16-Bit, 1.25-GSPS Digital-to-Analog Converter Evaluation Module," Texas Instruments, [Online]. Available: <https://www.ti.com/tool/DAC34H84EVM>. [Accessed 2021].
- [56] G. H. HAERTLING and C. E. LAND, "Hot-Pressed (Pb,La)(Zr,Ti)O<sub>3</sub> Ferroelectric Ceramics for Electrooptic Applications," *Journal of the American Ceramic Society*, vol. 54, pp. 1-11, 1971.



- [57] "PLZT photonics," [Online]. Available: <http://epiphotonics.com/products.html>. [Accessed 2021].
- [58] K. Nashimoto, "Nano-Second Speed PLZT Waveguide Switches and filters - EpiPhotonics White Paper," 2011. [Online]. Available: [http://epiphotonics.com/pdfs/PLZT%20Technology\\_EpiPhotonics.pdf](http://epiphotonics.com/pdfs/PLZT%20Technology_EpiPhotonics.pdf). [Accessed 2021].
- [59] XILINX, "7 Series FPGAs Data Sheet: Overview," 2020. [Online]. Available: [https://www.xilinx.com/content/dam/xilinx/support/documentation/data\\_sheets/ds180\\_7Series\\_Overview.pdf](https://www.xilinx.com/content/dam/xilinx/support/documentation/data_sheets/ds180_7Series_Overview.pdf).
- [60] FMC-ADC-ADAPTER, "High Speed ADC to FMC (Xilinx) Header Adapter Card," [Online]. Available: <https://www.ti.com/tool/FMC-ADC-ADAPTER>. [Accessed 2021].
- [61] P. Bakopoulos, K. Konstantinos, G. Landi, M. Aziz, E. Zahavi, D. Gallico, R. Pitwon, K. Tokas, I. Patronas, M. Capitani, C. Spatharakis, K. Yiannopoulos, K. Kontodimas, I. Lazarou, P. Wieder, D. I. Reisis, E. Varvarigos, M. Biancani and H. Avramopoulos, "NEPHELE: An End-to-End Scalable and Dynamically Reconfigurable Optical Architecture for Application-Aware SDN Cloud Data Centers," *IEEE Communications Magazine*, vol. 56, no. 2, pp. 178-188, Feb. 2018.
- [62] H.-s. F.-o. S. Solid-state, "Nanona High Speed & Low Loss Optical Switch," [Online]. Available: <https://www.bostonati.com/products/PI-FOS.pdf>. [Accessed 2021].
- [63] KAIAM, "80-Channel 50GHz Athermal AWG," [Online]. Available: <https://www.electronicdatasheets.com/download/520cb4f2e34e249d65df02c4.pdf?format=pdf>. [Accessed 2021].
- [64] I. HUSSEIN, "Application Note AN-2095, Controlling the S7500 CW Tunable Laser [Online]," [Online]. Available: <http://www.finisar.com/sites/default/files/pdf/AN-2095-ControllingtheS7500CWTunableLaser-RevA.pdf>.
- [65] Cisco, "Cisco Global Cloud Index: Forecast and Methodology," 2012-2017.
- [66]

- [67] "Cisco, "Cisco Annual Internet Report (2018–2023)", white paper,," [Online]. Available: <https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/white-paper-c11-741490.html> .
- [68] K. Tokas, C. Spatharakis, I. Patronas, P. Bakopoulos, G. Landi, K. Christodoulopoulos, M. Capitani, A. Kyriakos, M. Aziz, R. Pitwon, D. Gallico, D. Reisis, E. Varvarigos, E. Zahavi and H. Avramopoulos, "Real Time Demonstration of an End-to-End Optical Datacenter Network with Dynamic Bandwidth Allocation," in *2018 European Conference on Optical Communication (ECOC)*, Rome, 2018.
- [69] C. Spatharakis, K. Tokas, I. Patronas, P. Bakopoulos, D. Reisis and H. Avramopoulos, "NEPHELE: Vertical Integration and Real-Time Demonstration of an Optical Datacenter Network," in *2018 20th International Conference on Transparent Optical Networks (ICTON)*, Bucharest, 2018.
- [70] P. Bakopoulos, K. Tokas, C. Spatharakis, I. Patronas, G. Landi, K. Christodoulopoulos, M. Capitani, A. Kyriakos, M. Aziz, D. Reisis, E. Varvarigos, E. Zahavi and H. Avramopoulos, "Optical datacenter network employing slotted (TDMA) operation for dynamic resource allocation," in *SPIE OPTO, 2018*, San Francisco, California, United States, 2018.

# Λίστα Δημοσιεύσεων

## *Δημοσιεύσεις σε διεθνή έγκριτα περιοδικά με κρίση:*

1. **Konstantinos Tokas**, Giannis Patronas, Christos Spatharakis, Paraskevas Bakopoulos, Angelos Kyriakos, Giada Landi, Eitan Zahavi, Kostas Christodoulopoulos, Muzzamil Aziz, Richard Pitwon, "End-to-End Real-Time Demonstration of the Slotted, SDN-Controlled NEPHELE Optical Datacenter Network", MDPI Photonics, Multidisciplinary Digital Publishing Institute, 2020
2. Paraskevas Bakopoulos, Konstantinos Christodoulopoulos, Giada Landi, Muzzamil Aziz, Eitan Zahavi, Domenico Gallico, Richard Pitwon, **Konstantinos Tokas**, Ioannis Patronas, Marco Capitani, "NEPHELE: An end-to-end scalable and dynamically reconfigurable optical architecture for application-aware SDN cloud data centers", IEEE Communications Magazine, 178-188, 2018
3. Giannis Giannoulis, **Konstantinos Tokas**, Giannis Pouloupoulos, Giannis Kanakis, Panagiotis Toumasis, Konstantina Kanta, Dimitrios Apostolopoulos, Hercules Avramopoulos, "Integrated Photonic Filters in Support of Converged 5G Mobile Fronthaul & Midhaul Transport Layers", Fiber and Integrated Optics, Taylor & Francis, 2019
4. K. Kanta, P. Toumasis, **K. Tokas**, Stratakos I, Papatheofanous EA, Giannoulis G, Mesogiti I, Theodoropoulou E, Lyberopoulos G, Lentaris G, Apostolopoulos D, Reisis D, Soudris D, Avramopoulos H., "Demonstration of a Hybrid Analog–Digital Transport System Architecture for 5G and Beyond Networks", Applied Sciences 12, no.4, 2022

## *Δημοσιεύσεις σε συνέδρια (με κρίση):*

1. **Konstantinos Tokas**, Christos Spatharakis, Ioannis Patronas, Paraskevas Bakopoulos, Giada Landi, Konstantinos Christodoulopoulos, Marco Capitani,

- Angelos Kyriakos, Muzzamil Aziz, Richard Pitwon, “Real time demonstration of an end-to-end optical datacenter network with dynamic bandwidth allocation”, 2018 IEEE, European Conference on Optical Communication (ECOC), 2018
2. **Konstantinos Tokas**, Christos Spatharakis, Ioannis Kanakis, Nikolaos Iliadis, Paraskevas Bakopoulos, Hercules Avramopoulos, Ioannis Patronas, Nikolaos Liakopoulos, Dionysios Reisis, “A scalable optically-switched datacenter network with multicasting”, 2016 European Conference on Networks and Communications (EuCNC 2016),265-270, 2016
  3. **Konstantinos Tokas**, Ioannis Patronas, Christos Spatharakis, Dionysios Reisis, Paraskevas Bakopoulos, Hercules Avramopoulos, “Slotted TDMA and optically switched network for disaggregated datacenters”, IEEE 19th International Conference on Transparent Optical Networks (ICTON 2017), 2017
  4. Paraskevas Bakopoulos, **Konstantinos Tokas**, Christos Spatharakis, Ioannis Patronas, Giada Landi, Kostantinos Christodouloupoulos, Marco Capitani, Angelos Kyriakos, Muzzamil Aziz, Dionysis Reisis, “Optical datacenter network employing slotted (TDMA) operation for dynamic resource allocation”, Optical Interconnects XVIII, International Society for Optics and Photonics 2018
  5. Christos Spatharakis, **Konstantinos Tokas**, Ioannis Patronas, Paraskevas Bakopoulos, Dionysios Reisis, Hercules Avramopoulos, “NEPHELE: Vertical Integration and Real-Time Demonstration of an Optical Datacenter Network”, IEEE 20th International Conference on Transparent Optical Networks (ICTON 2018), 2018
  6. Paraskevas Bakopoulos, **Konstantinos Tokas**, Christos Spatharakis, Hercules Avramopoulos, “Slotted optical datacenter network with sub-wavelength resource allocation”, 2017 IEEE Photonics Society Summer Topical Meeting Series, 2017
  7. Paraskevas Bakopoulos, Stefanos Dris, Nikolaos Argyris, **Konstantinos Tokas**, Hercules Avramopoulos, “120 Gb/s PAM-8 and 80 Gb/s PAM-4 optical interconnect with a sub-volt driven EAM”, Asia Communications and Photonics Conference, Optical Society of America 2015

8. Giannis Giannoulis, Panagiotis Toumasis, Konstantina Kanta, Giannis Pouloupoulos, **Konstantinos Tokas**, Dimitrios Apostolopoulos, Hercules Avramopoulos, “Integrated Photonic Filters for Flexible Analog-based Mobile Fronthauling”, International Workshop on Fiber Optics in Access Networks (FOAN 2019), IEEE, 2019
9. Dimitris Zavitsanos, Argiris Ntanos, Panagiotis Toumasis, Adam Raptakis, **Konstantinos Tokas**, Konstantina Kanta, Christos Kouloumentas, Giannis Giannoulis, Hercules Avramopoulos, “On the Impact of Center Frequency Drifts on QKD Performance in WDM-based Nodes”, European Quantum Electronics Conference, Optical Society of America, 2021
10. P. Toumasis, K. Kanta, **K. Tokas**, I. Stratakos, E. A. Papatheofanous, G. Giannoulis, I. Mesogiti, E. Theodoropoulou, G. Lyberopoulos, G. Lentari, D. Apostolopoulos, D. Reisis, D. Soudris, H. Avramopoulos, “Demonstration of FPGA-based A-IFoF/mmWave transceiver integration in mobile infrastructure for beyond 5G transport”, 2021 European Conference on Optical Communication (ECOC), 2021
11. Francesca Moscatelli, Giada Landi, Agapi Mesodiakaki, Marios Gatzianas, George Kalfas, Anna Tzanakaki, Markos Anastasopoulos, Giannis Giannoulis, **Konstantinos Tokas**, Dimitris Apostolopoulos, Hercules Avramopoulos, Dimitra Simeonidou, Reza Nejabati, “5G-COMLETE: Service-Driven Slice Management over Heterogeneous 5G Infrastructures”, 2021 IEEE 26th International Workshop on Computer Aided Modeling and Design of Communication Links and Networks (CAMAD), 2021
12. Dimitrios Klonidis, Dimitris Apostolopoulos, Georgios P. Katsikas, Giannis Giannoulis, Konstantina Kanta, **Kostas Tokas**, Thanos Xirofotos, Raul Muñoz, Francesca Moscatelli, Guy Torfs, Christos Vagionas, David Larrabeiti Lopez, Zhongxia Simon He, Janez Sterle, Dotan Levi, George Lyberopoulos, Victor Lopez Alvarez, Eleni Trouva, Yigal Leiba, Xavier Vilajosana, J. Carles Terés Casals, Hercules Avramopoulos, “Int5Gent: An integrated end-to-end system platform for verticals and data plane solutions beyond 5G”, 2021 Joint European Conference on Networks and Communications & 6G Summit (EuCNC/6G Summit), 2021

13. Panagiotis Toumasis, Konstantina Kanta, **Kostas Tokas**, Giannis Giannoulis, Dimitris Apostolopoulos, Hercules Avramopoulos, “A Dynamically Reconfigurable Optical Switching Node for Hybrid Analog/Digital RoF Transport”, 2022 Optical Fiber Communications Conference and Exhibition (OFC), 2022
14. Konstantina Kanta, Panagiotis Toumasis, Giannis Giannoulis, Ioannis Stratakos, George Lentaris, Elissaios-Alexios Papatheofanous, Ioanna Mesogiti, Eleni Theodoropoulou, Aristotelis Margaritis, Dimitris Syrivelis, Evrydiki Kyriazi, George Brestas, **Kostas Tokas**, Nikos Argyris, Chris Vagionas, Ronis Maximidis, Paraskevas Bakopoulos, Agapi Mesodiakaki, Marios Gkatzianas, Georgios Kalfas, Kostas Tsagkaris, Nikos Pleros, Dionysios Reisis, George Lyberopoulos, Dimitrios Apostolopoulos, Dimitrios Soudris, Hercules Avramopoulos, “End-to-End Demonstration of an SDN-Reconfigurable, FPGA-Based TxRx Interface for Analog-IFoF/MmWave X-Haul”, 2022 IEEE, European Conference on Optical Communication (ECOC), ECOC 2022

# Λίστα Σχημάτων

Εικόνα 1: Μια κλασική fat-tree τοπολογία [12] .....	19
Εικόνα 2: Fat-tree δίκτυο κατασκευασμένο από μεταγωγείς σε μια τοπολογία Clos. ....	20
Εικόνα 3: Σχηματικό της προτεινόμενης υβριδικής αρχιτεκτονικής.....	27
Εικόνα 4: Επέκταση του δικτύου με την προσθήκη επιπλέον οπτικών επιπέδων.....	28
Εικόνα 5: Επισκόπηση του data plane της αρχιτεκτονικής δικτύου.....	31
Εικόνα 6: Η αρχική αρχιτεκτονική του Hybrid ToR switch.....	34
Εικόνα 7: Γενικό διάγραμμα του οπτικού τμήματος του ToR switch.....	35
Εικόνα 8: Απλοποιημένη εκδοχή του οπτικού τμήματος του ToR switch.....	35
Εικόνα 9: Απλοποιημένη υλοποίηση του ηλεκτρικού τμήματος του ToR switch .....	36
Εικόνα 10: Σχηματική απεικόνιση του επιπέδου ελέγχου της προτεινόμενης αρχιτεκτονικής .....	38
Εικόνα 11: Φωτογραφίες του μίνι datacenter κατά την περίοδο της προετοιμασίας. (α) Ο διακόπτης pod, συμπεριλαμβανομένου ενός WSS. (β) Τα γρήγορα λείζερ που χρησιμοποιούνται για τον πομπό. (γ) Τα FPGA extenders του ToR διακόπτη.....	39
Εικόνα 12: Διάγραμμα του πομπού της προτεινόμενης υβριδικής αρχιτεκτονικής. ....	40
Εικόνα 13: α) Σχηματική διάταξη της δομής λείζερ MG-Y που χρησιμοποιείται στην προτεινόμενη υβριδική αρχιτεκτονική μεταγωγής, β) (πάνω) δύο διαφορετικές χτένες συχνότητας με ελαφρώς διαφορετικό εύρος ελεύθερου φάσματος (FSR) που προέρχεται από τα διαφορετικά μήκη υπο-κοιλότητας, (κάτω) συνολική απόκριση συχνότητας του λείζερ με βάση το φαινόμενο Vernier.....	42
Εικόνα 14: Συχνότητα εκπομπής λείζερ MG-Y που απεικονίζεται σε συνάρτηση με τα αριστερά και δεξιά ρεύματα ανακλαστήρα. Μια ασυνεχής αλλαγή συχνότητας συμβαίνει στα όρια που υποδεικνύονται από τα μαύρα περιγράμματα [46].....	43
Εικόνα 15: (α) Το σχέδιο δύο διαστάσεων του κυκλώματος της πλακέτας (PCB), (β-γ) Απεικόνιση τριών διαστάσεων (3D) της πλακέτας με την χρήση σχεδιαστικού προγράμματος. ....	43
Εικόνα 16: Προσαρμοζόμενο λείζερ στο PCB. Στη φωτογραφία είναι ορατές οι γραμμές RF υψηλής ταχύτητας αλλά και οι γραμμές χαμηλής ταχύτητας που τερματίζουν σε SMA υποδοχείς.....	44
Εικόνα 17: Σχεδιαγράμματα α) Συνδέσεων των ρευμάτων στο Finisar S7500 λείζερ και β)Εσωτερικών δομών της συσκευής.....	44
Εικόνα 18: Πειραματική διάταξη που εφαρμόστηκε για τη ρύθμιση συντονισμού του μήκους κύματος και τη σταθερότητα ισχύος του λείζερ S7500. ....	46

Εικόνα 19: Διάγραμμα που απεικονίζει την μεταβολή του μήκους κύματος σε σχέση με την μεταβολή του ρεύματος στο τμήμα της φάσης του λέιζερ. Οι τιμές στα τμήματα των ανακλαστήρων του λέιζερ έχουν ρυθμιστεί σταθερά στις τιμές 8.56 mA (αριστερός ανακλαστήρας) και 9.93 mA (δεξιός ανακλαστήρας) [52].	51
Εικόνα 20: Μεθοδολογία για τον προσδιορισμό του χρόνου που απαιτείται για τον συντονισμό του μήκους κύματος του λέιζερ (tuning time)	51
Εικόνα 21: α) Το σήμα για τον δυναμικό χαρακτηρισμό του προσαρμοζόμενου λέιζερ: Κορυφή: ηλεκτρικό σήμα οδήγησης στον ανακλαστήρα. μεσαίο: οπτικό σήμα μετά το φιλτράρισμα στο $\lambda_1$ ; κάτω: οπτικό σήμα μετά το φιλτράρισμα στο $\lambda_2$ . β) Σχηματική αναπαράσταση για τον προσδιορισμό του χρόνου συντονισμού του λέιζερ	53
Εικόνα 22: Η πειραματική διάταξη που χρησιμοποιήθηκε για το δυναμικό χαρακτηρισμό του λέιζερ.	55
Εικόνα 23: Το εύρος ζώνης 3 dB του οπτικού ζωνοπερατού φίλτρου (Optical bandpass filter - OBF) με μεταβλητό μήκος κύματος και εύρος ζώνης, όπως απεικονίζεται σε ένα οπτικό φασματογράφο.	56
Εικόνα 24: Κοντινή απεικόνιση της μετάβασης του λέιζερ διαμορφωμένο στα 85Mb/s. Το λέιζερ εναλλάσσει την έξοδο από την κατάσταση $\lambda_1$ στα 1527,99nm (μπλε ίχνος, πάνω) στην κατάσταση $\lambda_2$ στα 1546,52nm (κόκκινο ίχνος, κάτω). Το πράσινο ίχνος δείχνει το ηλεκτρικό μοτίβο που οδηγεί τον αριστερό ανακλαστήρα.	57
Εικόνα 25: Μετρήσεις bit-error-rate (BER) σε σχέση με την λαμβανόμενη οπτική ισχύ στον δέκτη για δύο εναλλασσόμενα μήκη κύματος υπό δυναμικό έλεγχο του λέιζερ στα 85Mb/s.	58
Εικόνα 26: Η πειραματική διάταξη που χρησιμοποιήθηκε για την αξιολόγηση του προσαρμοζόμενου πομπού στα 10Gb/s.	59
Εικόνα 27: Στιγμιότυπο από το διάγραμμα ματιού (eye-diagram) του προσαρμοζόμενου πομπού στα 10Gb/s	60
Εικόνα 28: Καμπύλες BER σε σχέση με την λαμβανόμενη ισχύ για διαμόρφωση 10Gb/s για μήκη κύματος 1529.55 nm (ch.80), 1561.01 nm (ch.1) και 1554.94 nm (ch.16) που παράγονται από τον συντονιζόμενο πομπό.	60
Εικόνα 29: Η δομή των πακέτων της αρχιτεκτονικής του προτεινόμενου υβριδικού δικτύου	61
Εικόνα 30: α) Σήμα ελέγχου του δεξιού ανακλαστήρα του λέιζερ, β) Η έξοδος του προσαρμοζόμενου λέιζερ που εναλλάσσεται μεταξύ δύο μηκών κύματος ανάλογα με το σήμα ελέγχου που φαίνεται στο α), γ) διαμορφωμένα πακέτα σε ρυθμό 10Gb/s που παράγονται από τον προσαρμοζόμενο πομπό.	62
Εικόνα 31: Η πειραματική διάταξη για την δημιουργία και αξιολόγηση των 10Gb/s πακέτων της υβριδικής αρχιτεκτονικής.	63
Εικόνα 32: Περιοδικό μοτίβο δεδομένων στα 10Gb/s που παράγονται από τον PPG, προσομοιάζοντας τα πακέτα και το χρόνο παύσης της υβριδικής αρχιτεκτονικής.	64
Εικόνα 33: α) Τα πακέτα δεδομένων σε ρυθμό 10 Gb/s που παράγονται από τον PPG, β) το σήμα «κλειδώματος» που προέρχεται από τον PPG και χρησιμεύει ως σήμα συγχρονισμού για την προγραμματιζόμενη γεννήτρια σημάτων καθυστέρησης (SRS), γ) Το ηλεκτρικό	



σήμα δύο σταθμών που παράγεται από την SRS και χρησιμεύει στον έλεγχο του δεξιού ανακλαστήρα του λέιζερ, ώστε να εκπέμπει στα μήκη κύματος $\lambda_1=1535.822$ nm and $\lambda_2=1541.75$ nm.....	64
Εικόνα 34: Τα ίχνη που απεικονίζονται στον παλμογράφο πραγματικού χρόνου τα οποία δείχνουν την ταυτόχρονη λήψη των πακέτων σε διαφορετικά μήκη κύματος (μπλέ και κόκκινα ίχνη). Η κυματομορφή που απεικονίζεται με κίτρινο χρώμα δείχνει το ηλεκτρικό σήμα που παράγεται από τον PPG και εγγράφεται στους οπτικούς φορείς από τον ηλεκτροοπτικό διαμορφωτή MZM. ....	65
Εικόνα 35: Απεικόνιση του BER ως συνάρτηση της λαμβανόμενης οπτικής ισχύος για κάθε ένα από τα δύο μήκη κύματος $\lambda_1=1535.822$ nm (πράσινα σημεία) και $\lambda_2=1541.75$ nm (κόκκινα σημεία).....	67
Εικόνα 36: Ιστογράμματα συνάρτησης πυκνότητας πιθανότητας (Probability Density Function - PDF) των ανιχνευόμενων συμβόλων για $\lambda_1 = 1541,75$ nm σε λαμβανόμενες οπτικές ισχύεις -17 dBm (πορτοκαλί ίχνος) και -6 dBm (γκρι ίχνος), υποδεικνύοντας βελτιωμένη ποιότητα σήματος για υψηλότερα επίπεδα οπτικής ισχύος. ....	68
Εικόνα 37: Το εσωτερικό διάγραμμα λειτουργίας του 4×4 PLZT διακόπτη, που αποτελείται από 4 στάδια μεταγωγής. ....	68
Εικόνα 38: Φωτογραφία από τον γρήγορο 4×4 PLZT οπτικό διακόπτη καθώς και το διάγραμμα των ηλεκτρικών συνδέσεων που ελέγχουν τα στάδια μεταγωγής.....	69
Εικόνα 39: Η πειραματική διάταξη που χρησιμοποιήθηκε για τη μέτρηση των απωλειών στην οπτική ισχύ, ύστερα από τη διέλευση του φωτός μέσα από τον διακόπτη. ....	70
Εικόνα 40: Φωτογραφία που δείχνει την πειραματική διάταξη που χρησιμοποιήθηκε για τον χαρακτηρισμό του γρήγορου 4×4 PLZT οπτικού διακόπτη στο ΕΦΕ ..... 70	70
Εικόνα 41: Η Πειραματική διάταξη που χρησιμοποιήθηκε για τις μετρήσεις του χρόνου μεταγωγής του οπτικού διακόπτη. ....	73
Εικόνα 42: Χρόνος μεταγωγής του 4x4 οπτικού διακόπτη PLZT. (α) Χρόνος ανόδου και (β) Χρόνος πτώσης. Οι χρόνοι ανόδου και πτώσης μετρήθηκαν στο 10% και 90% του ρεύματος εξόδου της φωτοδιόδου (μέτρηση 10-90).....	73
Εικόνα 43: Η πειραματική διάταξη για τη μέτρηση της παρεμβολής (crosstalk) ..... 74	74
Εικόνα 44: Μέτρηση οπτικής ισχύος (α) στην έξοδο 3, σε σύγκριση με (β) τις υπόλοιπες εξόδους.....	74
Εικόνα 45: Σχηματική εικόνα του pod switch και διασύνδεση με ToR switches ..... 75	75
Εικόνα 46: Φωτογραφία του 2×2 PLZT διακόπτη της Epi-Photonics ..... 76	76
Εικόνα 47: Φωτογραφία του 1×2 οπτικού διακόπτη της BATi..... 76	76
Εικόνα 48: Αρχιτεκτονική της μονάδας διακόπτη επιλογής μήκους κύματος (WSS) με βάση την προσέγγιση «αποπολυπλεξία, μεταγωγή και αποπολυπλεξία». .... 77	77
Εικόνα 49: Πειραματική διάταξη για την αξιολόγηση της δρομολόγησης εντός pod ..... 79	79
Εικόνα 50: (α) Τα πακέτα που εκπέμπονται από τον πομπό στα μήκη κύματος $\lambda_1$ (κίτρινη κυματομορφή) και $\lambda_2$ (πράσινη κυματομορφή). Η μπλε κυματομορφή αντιπροσωπεύει το σήμα ελέγχου TTL για τον 1×2 οπτικό διακόπτη καθορισμού. (β) Μεγεθυμένο στιγμιότυπο	

του (α) όπου ο χρόνος παύσης (~10 μs) μεταξύ των πακέτων σημειώνεται με τις διακεκομμένες γραμμές. ....	79
Εικόνα 51: α) Μέτρηση BER ως συνάρτηση της λαμβανόμενης οπτικής ισχύος για τα ch.15 και ch.30, β) αντίστοιχα ιστογράμματα της πυκνότητας πιθανότητας (PDF) των ανιχνευθέντων συμβόλων στις λαμβανόμενες οπτικές δυνάμεις -14 dBm (πορτοκαλί ίχνος) και -6 dBm (γκρι ίχνος). ....	80
Εικόνα 52: Πειραματική διάταξη για την αξιολόγηση της δρομολόγησης εκτός rod .....	81
Εικόνα 53: Πειραματική διάταξη για την αξιολόγηση της λειτουργικότητας του WSS στο inter-rod σενάριο. ....	82
Εικόνα 54: α) Τα πράσινα και κίτρινα ίχνη απεικονίζουν τα πακέτα που έχουν προορισμό μέσα στο rod ενώ τα μπλε και κόκκινα ίχνη είναι τα αντίστοιχα σήματα ελέγχου για τους διακόπτες PLZT μέσα στο WSS. β) το κίτρινο ίχνος αντιστοιχεί στην οπτική κίνηση μετά την έξοδο του WSS με προορισμό το δακτύλιο WDM. ....	82
Εικόνα 55: α) BER ως συνάρτηση της λαμβανόμενης οπτικής ισχύος για τα κανάλια 15 (λ <sub>2</sub> ) και 30 (λ <sub>1</sub> ). β) Ιστογράμματα συνάρτησης πυκνότητας πιθανότητας (PDF) των bits που ανιχνεύθηκαν για λαμβανόμενη οπτική ισχύ -19 dBm (πορτοκαλί ίχνος) και -12 dBm (γκρι ίχνος), υποδεικνύοντας βελτιωμένη ποιότητα σήματος για υψηλότερα επίπεδα οπτικής ισχύος. ....	83
Εικόνα 56: Πειραματική διάταξη για την αξιολόγηση και δοκιμή δύο διαδοχικών στοιχείων οπτικής μεταγωγής: ένας διακόπτης PLZT στο rod 1 και ένας διακόπτης ως μέρος του WSS στο rod 2. ....	84
Εικόνα 57: Το μπλε ίχνος δείχνει το σήμα ελέγχου για τον διακόπτη καθορισμού στο rod 1 και το κόκκινο αντιπροσωπεύει το σήμα οδήγησης για τον διακόπτη μέσα στη μονάδα του WSS του rod 2. Η πράσινη κυματομορφή απεικονίζει τα πακέτα που έχουν εγγραφεί στο λ <sub>2</sub> (κανάλι 30). (α) Δύο κίτρινα πακέτα στο λ <sub>1</sub> (κανάλι 15) προορίζονται για το rod 2. (β) Ένα κίτρινο πακέτο στο λ <sub>1</sub> (κανάλι 15) προωθείται στον οπτικό δακτύλιο WDM. ....	84
Εικόνα 58: α) BER ως συνάρτηση της λαμβανόμενης οπτικής ισχύος για τα κανάλια 15 (λ <sub>1</sub> ) και 30 (λ <sub>2</sub> ). β) Ιστογράμματα συνάρτησης πυκνότητας πιθανότητας (PDF) των bits που ανιχνεύθηκαν σε λαμβανόμενες οπτικές ισχύς -21 dBm (πορτοκαλί ίχνος) και -14 dBm (γκρι ίχνος), υποδεικνύοντας βελτιωμένη ποιότητα σήματος για υψηλότερα λαμβανόμενα επίπεδα οπτικής ισχύος. ....	85
Εικόνα 59: Το υποσύστημα του προσαρμοζόμενου πομπού στον ToR switch. ....	87
Εικόνα 60: 3D απεικόνιση του προσαρμοζόμενου πομπού όπως σχεδιάστηκε. ....	87
Εικόνα 61: Ο 10Gb/s ενισχυτής RF που χρησιμοποιείται στον προσαρμοζόμενο πομπό. ....	88
Εικόνα 62: 10 Gb/s Mach-Zehnder Modulator Οπτικός Διαμορφωτής (MZM) .....	89
Εικόνα 63: (α) Πειραματική διάταξη αξιολόγησης του συστήματος 10Gb/s RF driver – Modulator και (β) Το διάγραμμα ματιού στα 10Gb/s που αποδεικνύει την μετάδοση χωρίς σφάλματα. ....	89
Εικόνα 64: Φωτογραφία από το προσαρμοζόμενο λέιζερ Finisar S7500. ....	90
Εικόνα 65: Το σύστημα FPGA/DAC για τον έλεγχο του προσαρμοζόμενου λέιζερ, τον συγχρονισμό των συστημάτων και την παραγωγή των πακέτων της αρχιτεκτονικής. ....	90

Εικόνα 66: α)Φωτογραφία από την ενσωμάτωση του DAC με το FPGA μέσω του προσαρμογέα FMC, β) Παράδειγμα από το γραφικό περιβάλλον σε υπολογιστή για τον χειρισμό του DAC.....	92
Εικόνα 67: Οι συνδέσεις μεταξύ του FPGA και του DAC.....	93
Εικόνα 68: Η διεπαφή επικοινωνίας μεταξύ του FPGA και του DAC.....	94
Εικόνα 69: Ηλεκτρικά σήματα που παράγονται από το DAC. Παρατηρείται ότι η μετάβαση από μια στάθμη σε μια άλλη διαρκεί σχεδόν 10 ns. ....	101
Εικόνα 70: Η Πειραματική διάταξη που χρησιμοποιείται για τον προσδιορισμό της συνολικής ταχύτητας συντονισμού του προσαρμοζόμενου πομπού. ....	102
Εικόνα 71: Τα ηλεκτρικά σήματα που παράγονται από το DAC για τη ρύθμιση μήκους κύματος του λέιζερ μεταξύ $\lambda_i$ και $\lambda_j$ . Τα πορτοκαλί, μοβ και μπλε ίχνη αντιπροσωπεύουν τα ρεύματα που εγχέονται στον αριστερό ανακλαστήρα του λέιζερ ( $I_i$ και $I_j$ ), στον δεξιό ανακλαστήρα ( $r_i$ και $r_j$ ) και στο τμήμα φάσης ( $p_i$ και $p_j$ ) αντίστοιχα.....	102
Εικόνα 72: (α) Έξοδος του 1×80 AWG που αντιστοιχεί στο $\lambda_i$ και (β) έξοδος του 1×80 AWG που αντιστοιχεί σε $\lambda_j$ και (γ) Το συνδυασμένο σήμα όπως ανιχνεύεται από τη φωτοδίοδο και χρησιμοποιείται για τη μέτρηση του χρόνου μεταγωγής για κάθε ζεύγος μηκών κύματος. ....	103
Εικόνα 73: Η συχνότητα εκπομπής ενός λέιζερ MG-Y απεικονίζεται ως συνάρτηση των ρευμάτων του αριστερού και του δεξιού ανακλαστήρα [46].....	106
Εικόνα 74: Σχηματική διάταξη δύο διαδοχικών pod της προτεινόμενης αρχιτεκτονικής	108
Εικόνα 75: Επισκόπηση του απλοποιημένου σχηματικού για τις πειραματικές δοκιμές.	109
Εικόνα 76: Απλοποιημένη πειραματική διάταξη για τη δοκιμή της επικοινωνίας εντός και μεταξύ pod μέσω δυναμικού ελέγχου από FPGA των οπτικών υποσυστημάτων. ....	110
Εικόνα 77: Πειραματική διάταξη για την αξιολόγηση της ενσωμάτωσης ToR και pod switch σε ένα intra-pod σενάριο δρομολόγησης .....	111
Εικόνα 78: Στιγμιότυπα των πακέτων που είναι εγγεγραμμένα στα $\lambda_1$ και $\lambda_2$ και περνούν από τη διαδρομή intra-pod. Τα σημεία A και B υποδεικνύονται στην πειραματική διάταξη του σχήματος 77. ....	111
Εικόνα 79: Το BER ως συνάρτηση της λαμβανόμενης ισχύος για το intra-pod σενάριο	112
Εικόνα 80: Πειραματική διάταξη για την αξιολόγηση της ενσωμάτωσης ToR και pod switch σε ένα inter-pod σενάριο δρομολόγησης .....	113
Εικόνα 81: (A-E) Στιγμιότυπα των πακέτων που αντιστοιχούν ίδια σημεία (A)-(E) που επισημαίνονται στην Εικόνα 80.....	114
Εικόνα 82: Το BER ως συνάρτηση της λαμβανόμενης ισχύος για το inter-pod σενάριο	115
Εικόνα 83: Πειραματική διάταξη που χρησιμοποιείται για την αξιολόγηση του σεναρίου επικοινωνίας μεταξύ δυο pod με έμφαση στη λειτουργικότητα του WSS. ....	116
Εικόνα 84: Στιγμιότυπα των πακέτων που εισέρχονται και εξέρχονται από το WSS. Τα στιγμιότυπα A, B και Γ αντιπροσωπεύουν τα σημεία που υποδεικνύονται στην πειραματική διάταξη της εικόνας 83. ....	117

Εικόνα 85: Το BER ως συνάρτηση της λαμβανόμενης ισχύος για το inter-pod σενάριο με έμφαση στη λειτουργικότητα του WSS.....	118
Εικόνα 86: Επισκόπηση της πειραματικής διάταξης και των κύριων στοιχείων που χρησιμοποιήθηκαν για την ενσωμάτωση του 4×4 οπτικού διακόπτη στο ToR switch. ...	119
Εικόνα 87: Η πειραματική διάταξη που χρησιμοποιήθηκε για την αξιολόγηση του intra-ToR σεναρίου εξολοκλήρου με οπτικό τρόπο. ....	120
Εικόνα 88: Τα ψηφιακά πακέτα 10G/s όπως αυτά δρομολογούνται μέσω του 4×4 PLZT οπτικού διακόπτη και κατευθύνονται είτε μέσα στο ίδιο ToR (intra-ToR scenario) είτε σε διαφορετικό (intra/inter-pod scenarios). Τα σημεία A,B και C στο σχήμα αντιστοιχούν στα ίδια σημεία στην Εικόνα 88. ....	121
Εικόνα 89: Ο ρυθμός εσφαλμένων bit (BER) σε σχέση με την λαμβανόμενη οπτική ισχύ για το intra-ToR σενάριο. ....	122
Εικόνα 90: Πειραματική διάταξη για την επικοινωνία ToR εντός του ίδιου pod με εξολοκλήρου οπτικό τρόπο. Τα σημεία A,B,C και D που διακρίνονται στο σχήμα αντιστοιχίζονται με αυτά στην Εικόνα 91 και δείχνει τον διαχωρισμό των πακέτων ανά μήκος κύματος και φυσικά ανάλογα με τη δρομολόγηση των οπτικών διακοπών.....	123
Εικόνα 91: Απεικόνιση των πακέτων που δρομολογούνται μέσω του 4×4 PLZT διακόπτη αλλά και του 1×2 διακόπτη που καθορίζει intra/inter-pod επικοινωνία. Τα σημεία A,B,C και D που διακρίνονται στο σχήμα αντιστοιχίζονται με αυτά στην Εικόνα 90 και δείχνει τον διαχωρισμό των πακέτων ανά μήκος κύματος και φυσικά ανάλογα με τη δρομολόγηση των οπτικών διακοπών. ....	124
Εικόνα 92: Μέτρηση BER σε σχέση με την λαμβανόμενη οπτική ισχύ για το σενάριο που αφορά επικοινωνία εντός του ίδιου pod με εξολοκλήρου οπτικό τρόπο.....	125
Εικόνα 93: Φωτογραφία από την πειραματική διάταξη όπως αυτή αναπτύχθηκε στο Εργαστήριο Φωτονικών Επικοινωνιών (ΕΦΕ) στο ΕΜΠ. ....	126
Εικόνα 94: Ο διακόπτης επιλογής μήκους κύματος (Wavelength Selective Switch) που υλοποιήθηκε βασιζόμενος στην προσέγγιση «αποπολυπλεξία, μεταγωγή και πολυπλεξία» για την δυναμική επιλογή μήκους κύματος. ....	127
Εικόνα 95: Πειραματική διάταξη για την εξολοκλήρου οπτική επικοινωνία ToRs που βρίσκονται σε διαφορετικά pod. Τα σημεία A,B,C,D και E που διακρίνονται στο σχήμα αντιστοιχίζονται με αυτά στην Εικόνα 96 και δείχνει τον διαχωρισμό των πακέτων ανά μήκος κύματος και φυσικά ανάλογα με τη δρομολόγηση των οπτικών διακοπών.....	128
Εικόνα 96: Απεικόνιση των πακέτων που δρομολογούνται μέσω του 4×4 PLZT διακόπτη, του 1×2 διακόπτη που καθορίζει intra/inter-pod επικοινωνία και φυσικά μέσω του WSS. Τα σημεία A,B,C,D και E που διακρίνονται στο σχήμα αντιστοιχίζονται με αυτά στην Εικόνα 95 και δείχνει τον διαχωρισμό των πακέτων ανά μήκος κύματος και φυσικά ανάλογα με τη δρομολόγηση των οπτικών διακοπών.....	130
Εικόνα 97: Μέτρηση BER σε σχέση με την λαμβανόμενη οπτική ισχύ για το σενάριο που αφορά την οπτική επικοινωνία μεταξύ υπολογιστών που βρίσκονται σε διαφορετικά pod. ....	131
Εικόνα 98: Φωτογραφία από την πειραματική διάταξη για την intra-pod επικοινωνία όπως αυτή αναπτύχθηκε στο Εργαστήριο Φωτονικών Επικοινωνιών (ΕΦΕ) στο ΕΜΠ.....	131

Εικόνα 99: Σχεδιασμός του μεταλλικού αντιστατικού ραφιού που θα περιέχει τα οπτικά υποσυστήματα στο πρότυπο datacenter .....	132
Εικόνα 100: Το rack του πρότυπου datacenter, που φιλοξενεί τα ράφια .....	133
Εικόνα 101: Τα όμοια λείζερ των πομπών κατά τη διαδικασία χαρακτηρισμού τους. ....	134
Εικόνα 102: Ένα από τα ράφια που φιλοξενεί όλα τα υποσυστήματα του ToR switch... 134	
Εικόνα 103: Φωτογραφία κατά τη διαδικασία δοκιμών ενός ToR switch. ....	135
Εικόνα 104: Βαθμονόμηση του RF driver έτσι ώστε να επιτυγχάνεται η ίδια απόδοση μεταξύ όλων των πομπών ToR. ....	136
Εικόνα 105: Το ολοκληρωμένο rack του πρότυπου mini-datacenter, που περιέχει όλα τα υποσυστήματα.....	137
Εικόνα 106: Το διάγραμμα της αρχιτεκτονικής δομής του πρωτότυπου datacenter .....	138
Εικόνα 107: Η διάταξη όλων των υποσυστημάτων που χρησιμοποιήθηκαν κατά την επίδειξη του πρωτότυπου datacenter .....	139
Εικόνα 108: Λεπτομερής σχηματική απεικόνιση των συνδέσεων από το κατώτερο επίπεδο του server στο υψηλότερο επίπεδο του πομπού και του δέκτη μέσα σε κάθε ToR.....	140
Εικόνα 109: Μια «ημέρα» στη ζωή ενός πακέτου.....	142
Εικόνα 110: Η δομή του επιπέδου δεδομένων στο ToR 2 για το σενάριο intra-ToR .....	145
Εικόνα 111: (α) Ο διακομιστής E3 (10.2.2.1) στέλνει κίνηση δεδομένων στον διακομιστή E4 (10.2.2.129) φτάνοντας το εύρος ζώνης 1,3 GBit/s, (β) Με την απελευθέρωση περισσότερων θυρίδων, το εύρος ζώνης διπλασιάζεται. ....	146
Εικόνα 112: Η ολοκληρωμένη διάταξη του intra-pod σεναρίου. Τα πακέτα επικοινωνίας μεταξύ ToR 2 και ToR 3 ακολουθούν τη διαδρομή που επισημαίνεται με τα μπλε βέλη 147	
Εικόνα 113: Στιγμιότυπα των πακέτων που αποστέλλονται από τον διακομιστή E3 (10.2.2.1 ToR2) και λαμβάνονται από τον διακομιστή ToR3 E5 (10.2.3.1): (α) Σε αυτήν την περίπτωση το 25% της περιόδου προγραμματισμού (20 slots από τα 80 μιας περιόδου) καταλαμβάνεται για τη διαδρομή επικοινωνίας μεταξύ ToR2 και ToR3 επιτυγχάνοντας εύρος ζώνης 1,8 GBit/s. (β) Σε αυτό το μισό της περιόδου (40 από τα 80 slots) καταλαμβάνεται για τη διαδρομή επικοινωνίας μεταξύ ToR2 και ToR3 επιτυγχάνοντας εύρος ζώνης 3,6 Gb/s. (γ) Σε αυτήν την περίπτωση, το 87,5% της περιόδου προγραμματισμού (70 από τα 80 slots) καταλαμβάνεται για τη διαδρομή επικοινωνίας μεταξύ ToR2 και ToR3 επιτυγχάνοντας εύρος ζώνης 6,4 Gbit/s. ....	148
Εικόνα 114: Πειραματική διάταξη για το σενάριο inter-pod. Οι σκούρες και οι ανοιχτές μπλε γραμμές αποκαλύπτουν τις οπτικές διαδρομές από τον server E1 προς τους server E3 και E5 αντίστοιχα.....	150
Εικόνα 115: (α) Τα πακέτα που αποστέλλονται από τον server E1 (10.1.1.1) του ToR1 και λαμβάνονται από το server E3 (10.2.2.1) του ToR2. Σε αυτήν την περίπτωση, το 75% της περιόδου προγραμματισμού (60 από τα 80 slots) καταλαμβάνεται για τη διαδρομή επικοινωνίας μεταξύ ToR1 και ToR2 επιτυγχάνοντας εύρος ζώνης 5,12 Gb/s. ....	150
Εικόνα 116: Στιγμιότυπα των πακέτων που προέρχονται από το ToR1 και ελήφθησαν από στο ToR 2 (μπλε ίχνος) και στο ToR 3 (πράσινο ίχνος). Η συνολική κίνηση διαιρείται εξίσου στους δύο προορισμούς (20slots φτάνουν σε κάθε ToR). ....	152

Εικόνα 117: (α) Ο server E1 (10.1.1.1) στέλνει κίνηση δεδομένων (β) στον server E3 (10.2.2.1) και (γ) server E5 (10.2.2.129) επιτυγχάνοντας εύρος ζώνης σχεδόν 1,4 Gb/s σε κάθε server. ....	152
Εικόνα 118: (α) Η πειραματική διάταξη που χρησιμοποιείται για την αξιολόγηση των συνδυασμών Intra-pod και Inter-pod σεναρίων, (β) Το σχηματικών των εμπλεκόμενων διεπαφών του data-plane των ToR switches. ....	154
Εικόνα 119: (α) Συνδυασμένο σενάριο I. (β) Η επαναλαμβανόμενη ακολουθία πακέτων σε τρία μήκη κύματος που παράγεται από τον προσαρμοζόμενο Tx του ToR 1: οκτώ πακέτα στο $\lambda_4$ , τέσσερα στο $\lambda_2$ και τέσσερα στο $\lambda_3$ . ....	155
Εικόνα 120: Στιγμιότυπα των ληφθέντων πακέτων σε κάθε ToR. Το κόκκινο ίχνος αντιστοιχεί στον δέκτη του ToR 0, το μπλε ίχνος στον δέκτη ToR 2 και το πράσινο ίχνος στον δέκτη ToR 3. ....	156
Εικόνα 121: (α) Σχηματικό για το συνδυασμένο σενάριο II Intra-pod και Inter-pod, (β) Επαναλαμβανόμενη ακολουθία πακέτων που μεταδίδονται από το ToR 1 και (γ) από το ToR 2 ταυτόχρονα και συγχρονισμένα στα $\lambda_2$ και $\lambda_3$ . ....	157
Εικόνα 122: Στιγμιότυπα των ληφθέντων πακέτων στους δέκτες των ToR 2 (μπλε πακέτα) και ToR 3 (πράσινα πακέτα). ....	158
Εικόνα 123: (α) Σχηματικό για το συνδυασμένο σενάριο III, (β) Επαναλαμβανόμενη ακολουθία πακέτων που μεταδίδονται από το ToR 1 και (γ) από το ToR 2 ταυτόχρονα και συγχρονισμένα στα $\lambda_2$ , $\lambda_3$ και $\lambda_4$ . ....	159
Εικόνα 124: Στιγμιότυπα των ληφθέντων πακέτων στους δέκτες των ToR 0, 2 και 3. Το κόκκινο ίχνος αντιστοιχεί στον δέκτη ToR 0 (dummy ToR), το μπλε ίχνος στον δέκτη ToR 2 και το πράσινο ίχνος στον δέκτη ToR 3. ....	160
Εικόνα 125: (α) Το δίκτυο επικοινωνίας για intra-DC και inter-DC δοκιμές (β) Το εύρος ζώνης που επιτεύχθηκε μέσω του διαδικτύου και μετρήθηκε στα 4,91 Mbit/s. ....	161
Εικόνα 126: Σχηματικό της προτεινόμενης υβριδικής αρχιτεκτονικής. ....	162
Εικόνα 127: Επισκόπηση του απλοποιημένου σχηματικού για τις πειραματικές δοκιμές. ....	167
Εικόνα 128: Το διάγραμμα της αρχιτεκτονικής δομής του πρωτότυπου datacenter. ....	170

# Λίστα Πινάκων

Πίνακας 1: Διαστάσεις του επιπέδου δεδομένων της προτεινόμενης αρχιτεκτονικής δικτύου .....	29
Πίνακας 2: Υπολογισμένες τιμές της αντίστασης του θερμίστορ του λέιζερ για διαφορετικές τιμές θερμοκρασίας.....	45
Πίνακας 3: Πίνακας συντονισμού λέιζερ, που δείχνει τα αντίστοιχα ρεύματα για: αριστερό και δεξιό ανακλαστήρα και τμήμα φάσης, αντίστοιχα. Η θερμοκρασία μέτρησης ήταν στους 25°C και το ρεύμα και για τα τμήματα κέρδους και SOA ορίστηκε στα 40 mA. ....	47
Πίνακας 4: Οι απώλειες που εισάγονται σε κάθε οπτική διαδρομή μέσα στον 4×4 οπτικό διακόπτη.....	71
Πίνακας 5: Οι τυπικές προδιαγραφές του 4×4 οπτικού διακόπτη PLZT. ....	72
Πίνακας 6: Σχέση καναλιών εισόδου και εξόδου του 8×8 cyclic AWGR που χρησιμοποιήθηκε στο εσωτερικό ενός pod switch .....	78
Πίνακας 7: Χαρτογράφηση σημάτων A, B και C που ελέγχουν τα ρεύματα DAC, ανά μήκος κύματος εξόδου του ρυθμιζόμενου λέιζερ.....	95
Πίνακας 8: Ενημερωμένος πίνακας τιμών ρευμάτων συντονισμού του λέιζερ, που δείχνει τα αντίστοιχα ρεύματα για τα τμήματα φάσης, αριστερού και δεξιού ανακλαστήρα, αντίστοιχα. Η θερμοκρασία κατά τις μετρήσεις ήταν σταθερή στους 25°C και το ρεύμα τόσο για τα τμήματα GAIN όσο και για το SOA ήταν στα 40 mA.....	98
Πίνακας 9: Χρόνος εναλλαγής σε νανοδευτερόλεπτα (ns) από το κανάλι 51 (1550,116 nm) σε καθένα από τα υπόλοιπα 79 κανάλια της C μπάντας, σύμφωνα με τα κεντρικά μήκη κύματος που καθορίζονται από το πρότυπο ITU-T.....	104
Πίνακας 10: Χρόνος συντονισμού μήκους κύματος λέιζερ για κάθε τύπο μετάβασης....	105
Πίνακας 11: Μήκη κύματος και IP διευθύνσεις που αντιστοιχούν σε κάθε server σύμφωνα με το pod και το ToR .....	141

# Extended Abstract

Today, datacenters (DC) are the heart of our online applications and Internet of Things (IoT) services, handling vast amounts of the digital information. The continuous enrichment of these online services and applications opens new vistas in user experience and sparks demand in bandwidth and speed. More and more value-added digital services spanning from virtual reality (VR) and high definition (HD) video streaming to cloud storage and sensor networks, forming the IoT, are sprouted all over the globe, burdening with heavy digital load the internet hubs. On the 5G era, all these activities turn to become more and more bandwidth hungry since the users demand instant and on-the-go access at any time. As a result, the datacenter networks (DCNs) should be capable to provide ultra-high capacity interconnects for a huge number of nodes and hosts, low latency to fulfil time-critical services and high reliability performance to reduce service interruption-time. Therefore, this profound effect emerges as an impact on the datacenter operation, driving worldwide datacenter network IP traffic on a steep growth curve reaching 25% annually [1].

Since the traffic within a DCN is much higher than the incoming/outcoming traffic [i], DCNs are facing remarkable challenges regarding the resource utilization, scalability and management agility. To successfully handle this soaring demand and avoid any possible capacity crunch and the relentless increase of power consumption, DCN operators and equipment providers are struggling to upgrade the existing infrastructures. Current state-of-the-art intra-DCNs are based on electronic switches interconnected in fat-tree topology or folded-clos topologies using fibers, with electro-opto-electrical conversion at each node [ii]. However, fat-tree topologies tend to underutilize resources and at the same time require a multitude of cables, fibers and switches.

Optical switching technologies are gaining traction as a potential vehicle to address the abovementioned challenging requirements. Deployment of photonic components could offer network scalability, due to their inherent speed, energy efficiency and transparency to protocol and bitrate. Several proposals based on optical technologies [iii] have been introduced as effective solutions within DCNs, such as space switching (e.g., using micro-electro-mechanical systems — MEMS or semiconductor optical amplifiers — SOAs [iv,v]), wavelength switching (through combination of tunable lasers with arrayed waveguide grating routers — AWGRs [vi,vii]), or a combination thereof (e.g., using wavelength-selective switches — WSSs [viii]). One of the key challenges currently pertaining to optical datacenter networks is the combination of scalability and fast reconfigurability. To this end, several efforts promote the integration of optical switching into control and orchestration



frameworks, the so-called software-defined networking (SDN) [ix,x]. Indeed, SDN platforms in combination with orchestration algorithms, provide dynamicity and scalability in DCNs and enhance the benefits of the optical switches.

In this doctoral thesis a dynamic end-to-end optical network infrastructure for large-scale and disaggregated datacenters is proposed and presented [xi]. In this context, the proposed architecture combines optical switching benefits with SDN control and orchestration to beat current datacenter challenges. To achieve this and following a vertical development approach, the proposed architecture is expanding from the datacenter architecture to the overlaying control plane, in order to deliver a fully functional networking solution. This proposal brings two ambitious innovations:

The **data plane architecture**, leverages commercial-off-the-shelf (COTS) photonic components in combination with slotted time-division-multiple-access (TDMA) operation to enable dynamic and efficient sharing of resources [xii,xiii].

An **SDN orchestration and control framework** is responsible for the management of all the underlying data plane elements, extending the open-source SDN platforms with TDMA functionalities. From this angle, the orchestration framework is capable of dynamically assigning network resources directly at the optical layer [xiv]. Multiple algorithmic add-ons focusing on the fast resource allocation were developed and integrated to the SDN platform [xv].

The rest of this manuscript is organized as follows. Section 2 summarizes the proposed architecture; the data plane and the control plane. In Section 3, we introduce the demonstrator [xvi,xvii] assembly, that was built at the Photonics Communications Research Laboratory in the National Technical University of Athens, while its performance is assessed through several real-time and end-to-end communication scenarios presented in Section 4. Finally, Section 5 concludes the paper.

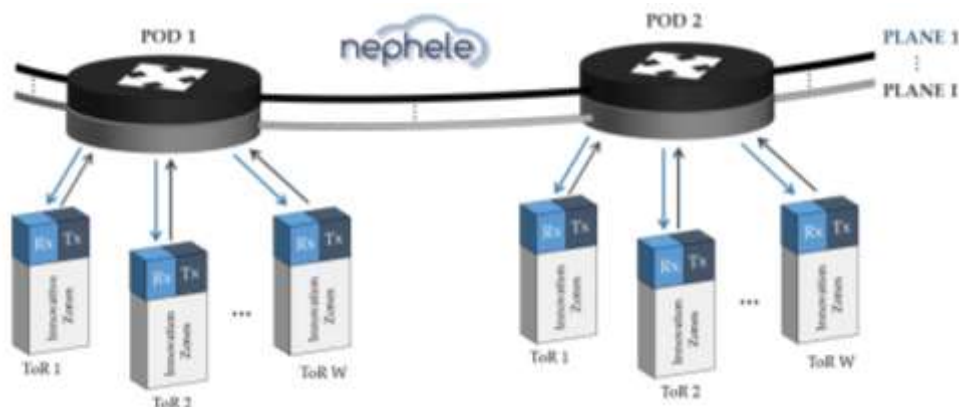
## **The proposed hybrid network Architecture**

The proper operation of the proposed DCN is based on an architecture where two tiers are collaborating and coexisting seamlessly: the data plane and the control plane. In this section we present an overview of the proposed architecture, highlighting the key innovations and functionalities of the network.

### *2.1 Data plane Overview*

The network adopts a flat and scalable topology that utilizes active and passive optical components to overcome the shortcomings of hierarchical topologies that are broadly used in electrical DCN architectures. As shown in Figure 1, the proposed architecture consists of

two layers of switches: the ToR (Top-of-Rack) and the POD switches. This way, the proposed topology serves efficiently both the north-south and east-west communication, which is an important advantage compared to conventional DC topologies. In addition, the required number of network modules scales linearly with the end-nodes, while the fat-tree network scales super-linearly requiring the addition of switches at all levels and the addition of extra levels once all ports are connected.



**Figure 1:** Overview of data plane network architecture

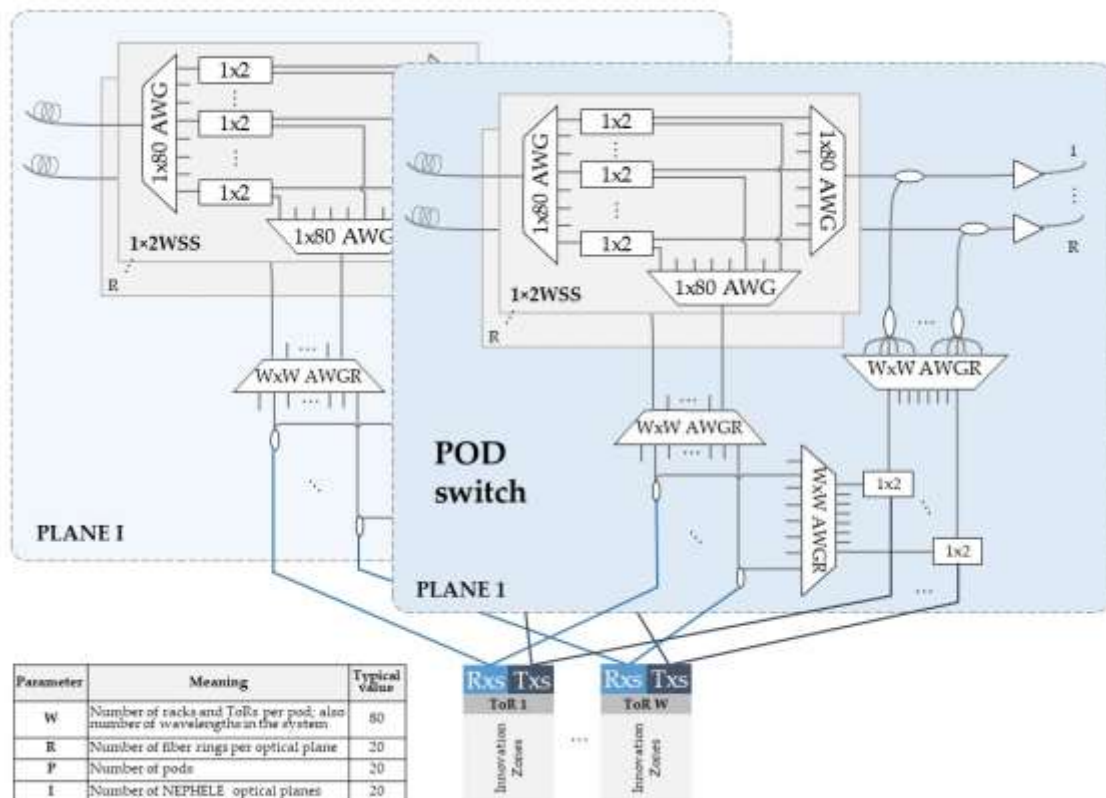
The proposed network architecture relies on PODs, which in a sense are small Datacenters, accommodating several racks. Each rack is regulated by a Top-of-Rack (ToR) switch and consists of several hosts (i.e. disaggregated storage and compute resources, hereafter called “innovation zones”). The ToRs are connected to the POD switch following the star topology structure. Network scalability is achieved by interconnecting multiple PODs in a Dense Wavelength Division Multiplexing (DWDM) multiple-fiber ring. Moreover, multiple parallel planes interconnecting the PODs, as it is shown in Figure 2, further scale the overall throughput of the network. The optical plane refers to the ensemble of a multi-fiber ring along with its corresponding POD and ToR interfaces. Each optical plane is connected to a different port of each ToR.

The proposed data plane operates in a slotted Time-Division-Multiple-Access (TDMA) manner. The slot duration is 200us with additional 10us as guard-time. The guard-time was defined by the response of the non-ideal DC-coupled electronics and the lock time of the FPGA receiver, while the slot duration was specified at 200us in order to provide 95% network utilization. Furthermore, slotted operation facilitates the dynamic resource allocation of the network, offering dynamic reconfiguration with sub-wavelength granularity. The scheduling (resource allocation) process is realized in a periodic manner. The time is divided in scheduling periods, with each period consisting of 80 slots (16ms). At the end of each scheduling period, the scheduler calculates the configuration of the network for each slot of the next scheduling period. The control plane and scheduling are

discussed in section 2.2. The following subsections provide an overview of the functionalities of the data plane elements that compose the proposed network.

### 2.1.1 The Pod switch (POD)

The pod switch (POD) distinguishes the *inter-* and *intra-pod* traffic and is based on two types of active optoelectronic photonic modules: a Wavelength Selective Switch (WSS) based on the “demultiplex, switch and multiplex” approach and a  $1 \times 2$  optical fast switch. A schematic of the POD switch is given in Figure 2. In the upstream direction (from the hosts to the optical network) the  $1 \times 2$  switches direct the traffic either to another ToR of the same POD (*intra-pod*) or to the WDM rings (*inter-pod*). In both cases,  $W \times W$  arrayed waveguide grating routers (AWGR) are used to route the signals to the appropriate destination. In the downstream direction (from the optical network to the hosts) the WSS drops wavelengths from the rings to the appropriate PODs. The wavelengths are routed to the destination ToR through AWGRs. The combination of these modules with several passive filtering photonic elements within the network, allows wavelength reuse among PODs, enabling network scalability beyond the typical wavelength count of DWDM systems. The configuration of the active components of the POD is decided by the SDN controller of the network (section 2.2) and was realized by FPGAs during the experiments described in the following sections.



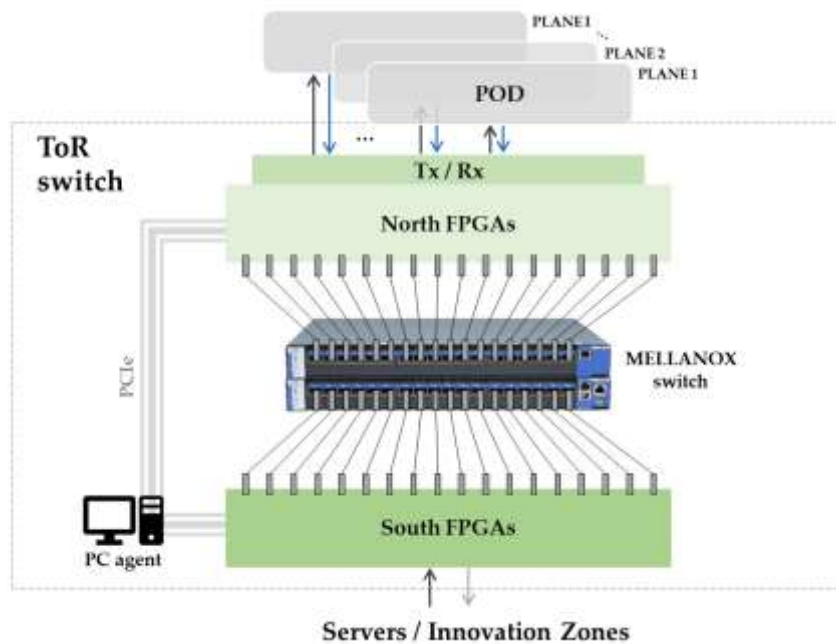
**Figure 2:** The pod switch schematic. Additional optical planes are scaling up the network

### 2.1.2 The Top-of-Rack switch (ToR)

Each Top-of-Rack switch (ToR) interconnects the hosts in the datacenter racks as well as to the higher network layer, handled by the PODs. To support the slotted operation of the optical switching fabric, we developed functionality extenders of a commercial electrical ToR switch (Figure 3). The extenders are developed on FPGA boards and they are placed on the “South” (between the electrical switch and the servers) and on the “North” (between the electrical switch and the optical network) of a commercial electrical switch (Mellanox SX1024). This way, in a future DC, the innovation zones and the applications running on them will remain transparent to the optical network restrictions, while the ToR switch will undertake their seamless integration with the slotted optical network. Note here that, the FPGAs across the proposed network are synchronized in time and thus the operations described in the following subsections are coordinated in sub-us scale.

#### 2.1.2.1 South-FPGA (S-FPGA) extender

The South FPGA resides between the servers and the electrical ToR switch. It receives the Ethernet frames generated from the servers in the innovation zones, parses the headers and stores them in VOQs (Virtual Output Queues) per destination ToR and input port. In order to efficiently utilize the available memory, the VOQs are dynamically created depending on the incoming traffic. The South FPGA forwards chunks of VLAN (Virtual Local Area Network) tagged Ethernet frames with the same ToR destination (i.e. from the same VOQ) to the Ethernet switch. The South FPGA has a bidirectional communication channel with the control plane. It notifies the SDN controller, and therefore the scheduler, about the status of its VOQs in a periodic manner. In the opposite direction, the SDN controller sends the following instructions: a) which VOQ will be emptied for each of the upcoming slots and b) a VLAN tag for each slot (which via the electrical ToR switch will define the outgoing port/plane of transmission on the optical network).



**Figure 3:** A schematic of the inside of a ToR

#### 2.1.2.2 Legacy electrical switch

In the electrical switch we use two different schemes of switching, depending on the direction of the traffic. The frames received from the South FPGA, and thus with direction from the servers to the optical network (upstream), are switched based on their VLAN tag. The VLAN tag has been inserted in the South FPGA according to the SDN controller's instructions to steer the chunks of ethernet frames to the appropriate plane. On the opposite direction, from the optical network to the servers (downstream), the switching is based on the destination MAC/IP address. This way, the frames that arrive at the ToR switch from the optical fabric are forwarded to the appropriate server/innovation zones.

#### 2.1.2.3 North FPGA(N-FPGA) extender

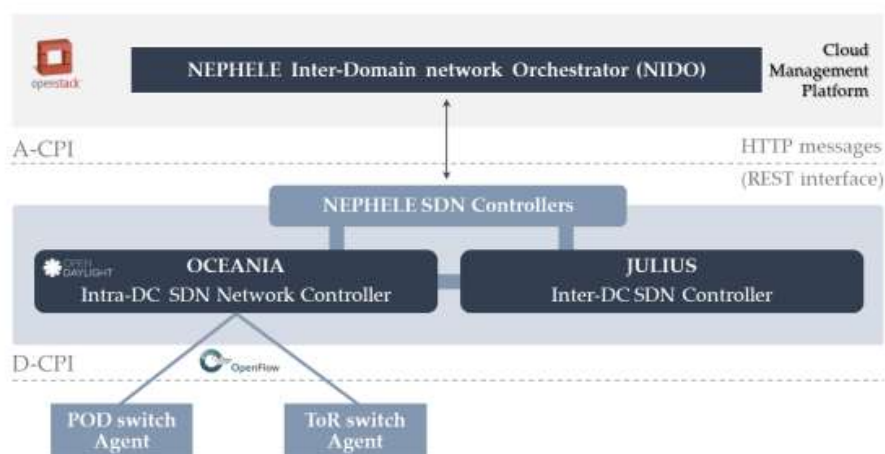
As described in the previous subsections, the North FPGA receives the chunks of ethernet frames from the electrical switch. Its role is to handle the interfacing with the optics and to realize the slotted operation. The North FPGA includes a slot-sized FIFO to fine tune the timing on which data are sent on the optical network. The chunks of Ethernet frames are encapsulated in the frame that the proposed architecture introduces: we add a ~8us long preamble to facilitate the CDR locking on the receiver, a delimiter and a short header with management fields. On the receiving side, the Ethernet frames are decapsulated from the frame and they are forwarded to the electrical switch.

The key building block of the Top-of-Rack (ToR) switch is the wavelength-tunable transmitter (Tx), which consists of an FPGA-controlled tunable laser, an MZI modulator and an RF driver [Error! Bookmark not defined.]. The tunable Tx is responsible for i

multiplexing the 10 Gb/s electrical data onto a selectable optical carrier (wavelength). The transmitter behaves in bursts and obeys the TDMA operation. The wavelength that each North FPGA uses in each slot is dictated by the SDN controller according to the destination ToR.

## 2.2 Control-Plane Overview

The control and orchestration framework is applied on both intra-DCN and inter-DC domain. This means that the control plane is based on a centralized inter-domain network orchestrator, named NIDO [xviii] (NEPHELE Inter-Domain network Orchestrator), that coordinates at multiple DCNs and the intermediate interconnection nodes to achieve end-to-end resource allocation [xix]. More specifically, NIDO orchestrates the actions of lower layer intra-domain SDN controllers, namely OCEANIA [xx] (OptiCal Electrical ApplicatioN Aware data centre network controller) for proposed-based DCNs and JULIUS [xxi] for inter-DC communication, as it is depicted in Figure 4. This hierarchical approach enables to change the intra-domain controller and/or the related network technology of any domain (DC, core, metro, access network).



**Figure 4:** The inter-domain Orchestration Architecture

The intra- DCN is controlled through the OCEANiA controller, an SDN controller that extends the open source OpenDaylight controller (Lithium version). Furthermore, OCEANiA works in-line with SDN applications, algorithms and OpenFlow protocol extensions to efficiently operate a proposed DCN. It caters for the dynamic establishment, re-configuration and tear-down of intra-DC connections by optimizing the allocation of the space (planes) and time (time-slots) resources. In the development of the prototype achieved the integration of the OCEANiA SDN controller with the data plane at the south-bound and the Inter-DC Orchestrator (NIDO) at the north-bound. The OCEANiA controller was demonstrated successfully in [xiv]. At the SDN application level, new resource allocation (scheduling) algorithms were implemented to enable an efficient and fully automated path

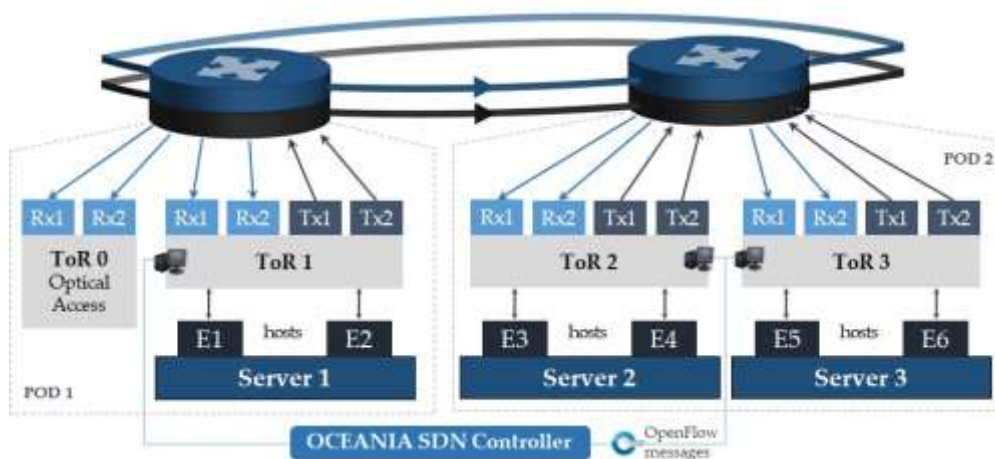
allocation. Extra effort was put to develop scheduling algorithms that would be efficient in large DCNs but also perform fast calculations to enable rapid network reconfiguration [xv,xxii].



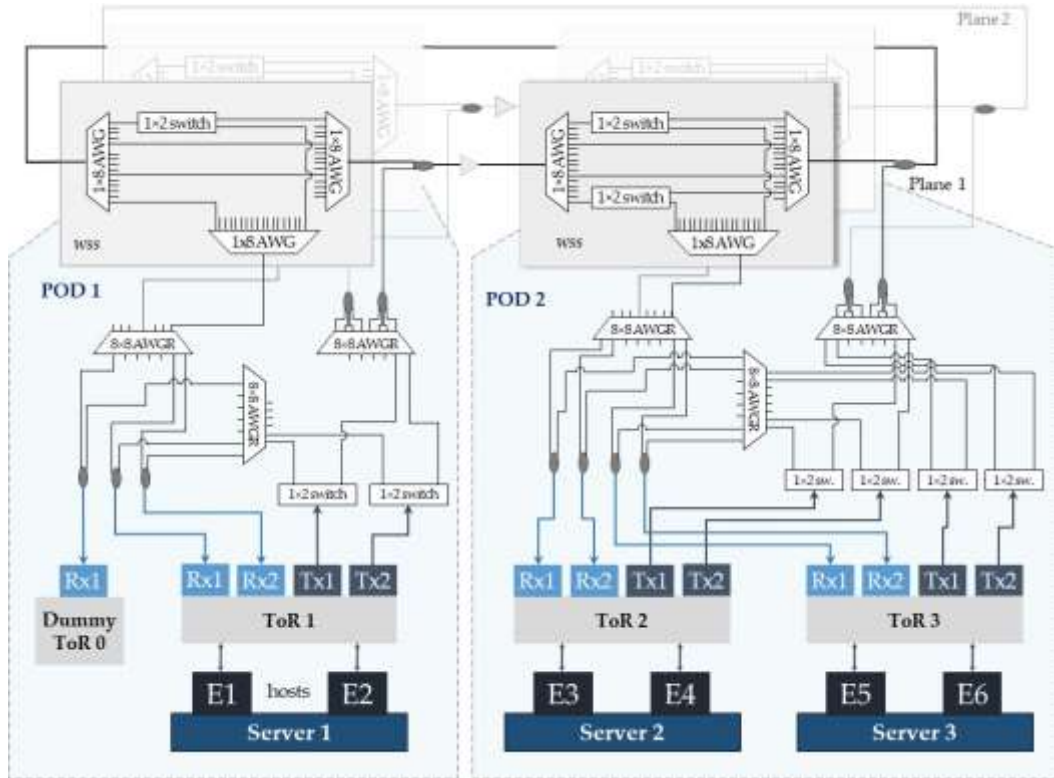
**Figure 5:** Photos of the demonstrator during the preparation period. (a) The POD switch, including a WSS with the “Demultiplex, switch and Multiplex” approach. (b) The fast tunable lasers used for the transmitter. (c) The FPGA boards that assemble the ToR switch

### 3. The Demonstrator Assembly

The demonstrator setup was built step-by-step at the Photonics Communications Research Laboratory in the National Technical University of Athens [xxiii]. Photos of the demonstrator during the preparation period are shown in Figure 5 and the implemented setup is graphically depicted in the schematic of Figure 6. In this setup two optical planes with two PODs are emulated. More specifically, POD 1 accommodates a server with two independent 10 Gb/s Ethernet interfaces/hosts (E1 and E2) through ToR 1 and a dummy ToR switch (ToR 0) for monitoring purposes. On the other hand, POD 2 handles two servers, each one accommodating two hosts through ToR 2 (E3 and E4) and ToR 3 (E5 and E6) respectively.



**Figure 6:** The optical network used for the real-time demonstration



**Figure 7:** The complete experimental setup used for the live demonstration

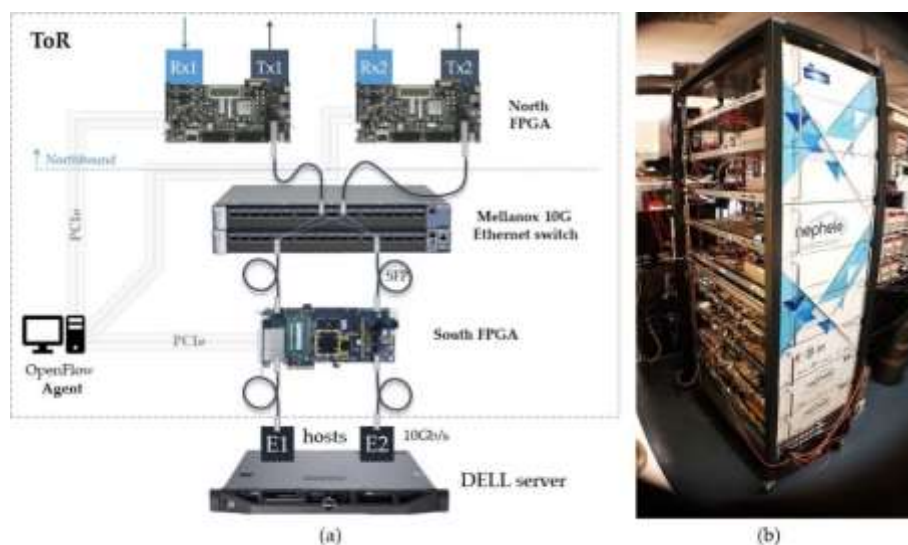
The demonstrator was assembled according to the architectural principles, in lab-scale dimensions for the purposes of the demonstration. Each of the ToR switches is equipped with two tunable transmitters (Tx) and equal number of optical receivers (Rx), as depicted in Figure 7. In this context, the demonstrator is composed of 3 fully equipped ToR switches, i.e. 6 transmitter and receiver assemblies that were developed, tested and used for the demo [xii,xiii,**Error! Bookmark not defined.**]. Finally, a partially functional receiver was connected to the “dummy” ToR 0, which is in POD 1 and it was used for displaying purposes.

Moreover, each POD switch employs two Wavelength Selective Switches (WSS) based on the “Demultiplex, switch and multiplex” approach [xvii]; one for each optical plane. Passive optical filtering elements (8×8 Arrayed Waveguide Grating Router AWGRs) and couplers are used in order to achieve all possible communication scenarios between the hosts through a combination of wavelength- and space-switching. Moreover, optical amplifiers (EDFAs) were used to amplify the optical signals in all optical light-paths between each POD and plane.

Figure 8 (a) shows a detailed schematic of the internal connections between the ToR switch and the hosts. Starting the description from the bottom to the top, three servers were used as hosts inside the racks, serving as the innovation zones. Each of the servers can produce 10Gb/s Ethernet traffic and it is connected to the South FPGA (S-FPGA) by means of two 10 Gbps SFP interfaces [xxiv]. The interfaces are independent and as a result they can



emulate different hosts. Indeed, server 1 located in POD 1, handled by ToR1, is connected to the S- FPGA by E1 and E2 Ethernet interfaces. After that, the traffic is forwarded to the commercial Ethernet switch, which sends the frames to the available North FPGA (N-FPGA) for transmission. The N-FPGAs control the tunable Tx and encapsulate the frames (200 $\mu$ s duration, including 10 $\mu$ s guard-time). In the Tx, the frames (in the electrical domain) are amplified in an RF driver and afterwards they are introduced to a Mach-Zehnder Modulator. The optical carrier is provided by a tunable laser assembled on a custom carrier, which is also fully controlled by the N-FPGA. The reception of the flows follows the opposite vertical direction.



**Figure 8:** (a) Detailed schematic of the connections from the lower level of the host to the higher level of transmitter and receiver inside each ToR. (b) Fully populated rack, containing the optical assemblies as well as the Ethernet switches and servers serving as innovation zones.

From the physical layer perspective, the communication between the servers in the network is achieved by using different wavelengths to route the traffic leveraging the combined effect of the fast tunable lasers and the AWGRs. On the other hand, regarding the network/protocol tier, IP addresses are assigned to each ethernet interface. Table 1 shows the wavelengths and IP addresses that correspond to each interface according to the hosting POD and ToR.

With respect to the control plane, the SDN controller and the scheduler of the DC are running on a remote machine that is connected to the testbed via a 1Gbps local area network. The SDN controller is connected with the SDN agents, which run on desktops that are also connected to the North and South FPGAs (through the PCIe interface). The control plane's instructions are calculated, transferred and enforced by the FPGAs in real time for the DC demonstration.

**Table 1:** Wavelengths and IP addresses that corresponds to each interface according to the hosting POD and ToR

Destination ToR	Located POD	Wavelength $\lambda$ (nm)	Interface	IP address
ToR 1	1	1546.917	E1	10.1.1.1
			E2	10.1.1.129
ToR 2	2	1547.715	E3	10.2.2.1
			E4	10.2.2.129
ToR 3	2	1548.515	E5	10.2.3.1
			E6	10.2.3.129
ToR 4	1	1549.315	Dummy ToR	

### 3.1 A “day” in the life of a Packet

The current subsection gives a high-level description of a packet’s journey in the proposed network, aiming to provide a better understanding of the end-to-end data transmission. Each end-host is assigned with a static IP (Internet Protocol) address using DHCP (Dynamic Host Configuration Protocol) or OpenFlow. Each rack is an IP subnet, thus the IP addresses of all the hosts in the rack begin with the same prefix (we use 24-bit prefix). The host address within the IP subnet is the index of the NIC (Network Interface Card) within the ToR.

Ethernet frames are transmitted from the NIC/host to the input ports of the ToR (ToR1 in Figure 9) through the S-FPGA. These frames carry the DMAC (Destination Media Access Control) address of the destination host. Static ARP (Address Resolution Protocol) can be used for mapping the destination host IP address to MAC address. The static ARP tables should be provided by the SDN controller. The frames sent from the hosts reach the ToR and are stored in buffers sorted by input port and destination ToR pair. These buffers for a specific input port are drawn in Figure 9 within the InP1 box. The schedule provided by the SDN controller and held by the ToR, is executed slot after slot. The SDN controller’s instructions for each slot dictate which buffer [Input Port, Destination ToR] to send at the current time, the wavelength to be used and through which Plane the frame will be routed. If there are frames in that buffer, they are sent to the ToR Ethernet switch after they are tagged with a VLAN tag that matches the destination Plane defined by the schedule. The ToR Ethernet switch forwarding inter-ToR traffic is statically configured using OpenFlow.

Each VLAN tag is mapped to a different “North Port” connected to a POD switch. In this way the incoming VLAN tagged frames to the Ethernet switch are forwarded to the correct Plane.

On their way to the tuneable laser that drives the POD switch on that Plane, the Ethernet frames are being stripped from the VLAN tag and are encapsulated in a frame. Note that the entire ToR switch cannot be classified as a VOQ switch since frames are being stored per input port per destination ToR switch (to consider this as a VOQ switch the output ports should be the ToRs and not the planes as in our case). The POD switch is also controlled by an OpenFlow agent and is performing its pre-programmed schedule. Each of the schedule slots defines the state of the fast switch and the WSS (for each colour). According to the current slot the POD1 is aware whether this frame concerns Inter-POD or Intra-POD traffic. In the case of Inter-POD traffic, the frame is routed via the 1x2 fast switch and the AWGR towards the optical ring.

In Figure 9 inter-POD communication is illustrated and POD 2 is instructed to drop the wavelength that carries the data from ToR1 in the slot that the communication takes place. The arriving frames are stripped of their header before they reach the Ethernet switch on the destination ToR2. The Ethernet switch in ToR2 is statically configured by OpenFlow to forward intra-ToR traffic using the DMAC that is incorporated in the Eth header. In this way the Ethernet frames are forwarded to the right port (NIC) depending on the DMAC. Optionally the forwarding could have been realized using the IP address of the destination NIC. The same forwarding rules of the Ethernet switch allow the forwarding of the intra-ToR traffic within ToR1.

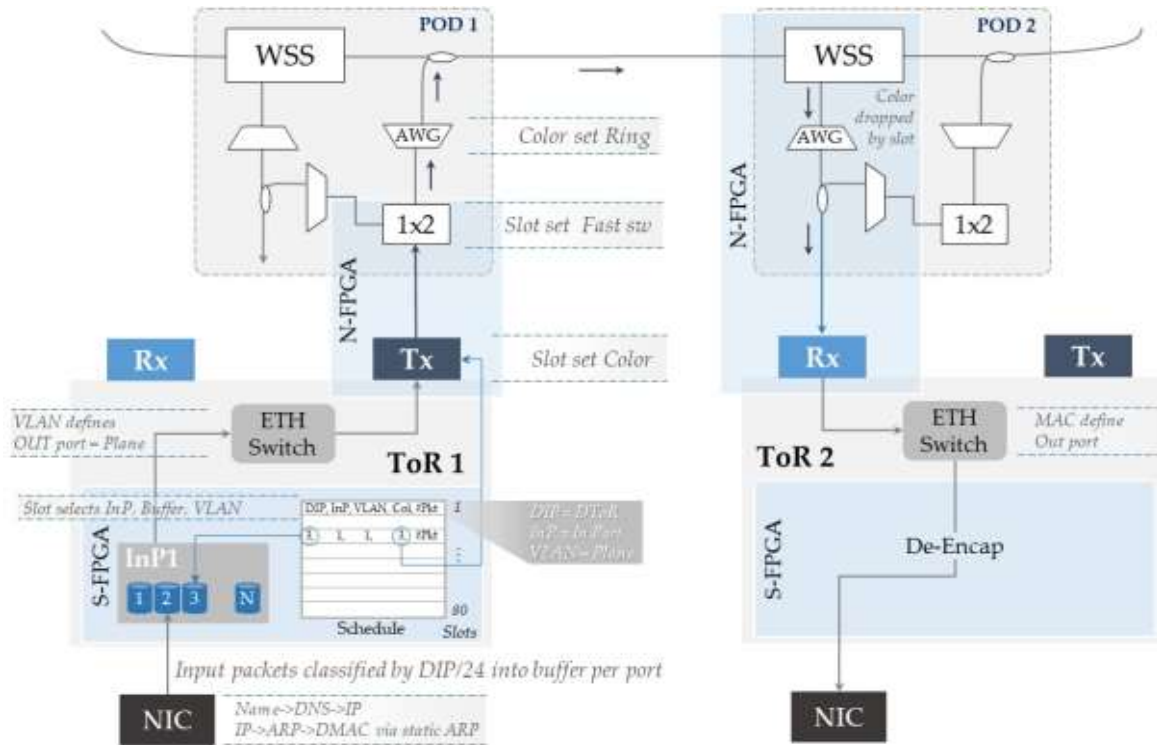


Figure 9: A “day” in the life of a Packet

#### 4. End-to-end Communication Scenarios and Real-Time Demo Results

In this section, we present the results of the end-to-end real-time operation of the NEPHELE demo DCN. More specifically, the *intra-* demo DCN communication scenarios are presented in subsection 4.1, while an *inter-* Demo DCN interconnection between the Athens demo and the Pisa testbed in Italy is presented in subsection 4.2.

##### 4.1. Intra-Datacenter Communication

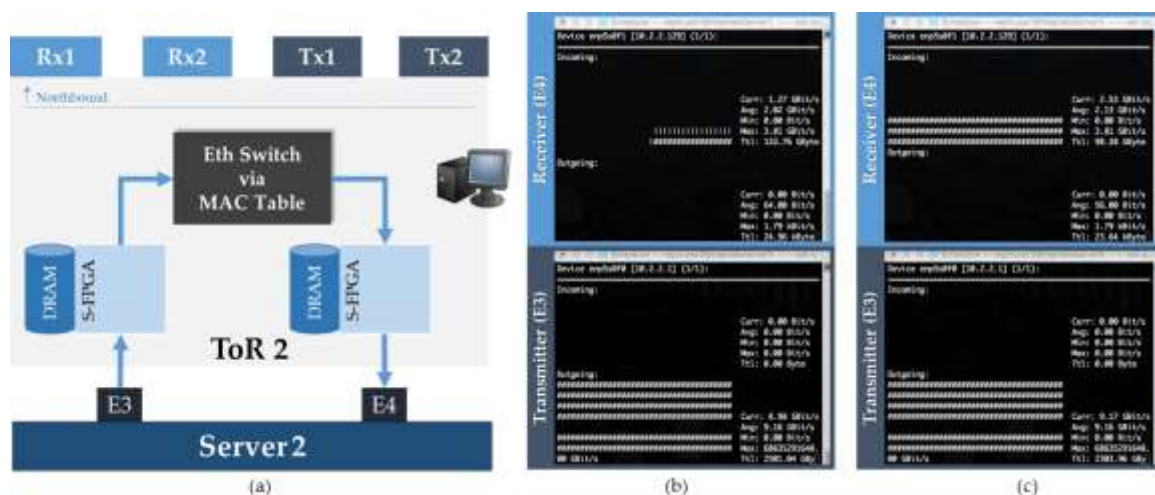
The intra-DC communication section consists of the communication scenarios that are taking place within the mini-DCN. Taking into account the proposed demonstrator of Figure 6, the scenarios are categorized as follows: *Intra-ToR*, *Intra-POD* and *Inter-POD* communication scenarios, depending on the location of the communicating hosts within the DCN.

###### 4.1.1 Intra-ToR scenario

The first scenario demonstrated was the *intra-ToR* traffic case. In this scenario, the Ethernet traffic remains inside the same rack and as a result the northbound part of the ToR switch is not involved. Figure 10(a) shows the block diagram of the demo setup and the blue arrows indicate the traffic flow paths; from E3 to E4 Ethernet interface within the ToR 2.

In Figure 10(b-c) the *nload* command is depicted running simultaneously in all the involved host-servers. In this way the incoming and outgoing traffic on each server is monitored.

More specifically, in Figure 10(b) the communication between server E3 and E4 is shown; server E3 (10.2.2.1) is sending data traffic to server E4 (10.2.2.129). The bandwidth achieved in this case is almost 1.3 GBit/s. In Figure 10(c) the traffic follows the same path but more slots are allocated by the scheduling engine and as a result the bandwidth is increased by a factor of almost two. In the described scenario no optical part of the POD switch is used. The traffic is handled by the south part of the ToR switch, since the involved hosts belong to the same rack.



**Figure 10:** (a) The data plane structure of the ToR 2 for the intra-ToR scenario, (b) Server E3 (10.2.2.1) is sending data traffic to server E4 (10.2.2.129) reaching 1.3GBit/s bandwidth, (c) By allocating more slots the bandwidth is doubled.

#### 4.1.2 Intra-Pod scenario

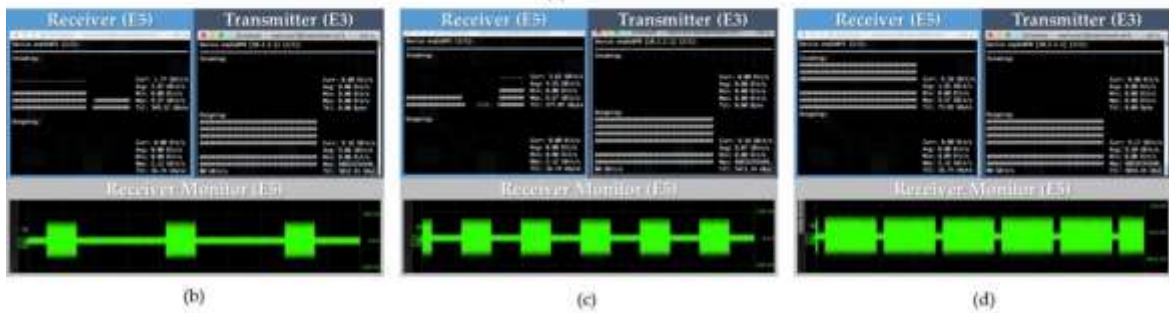
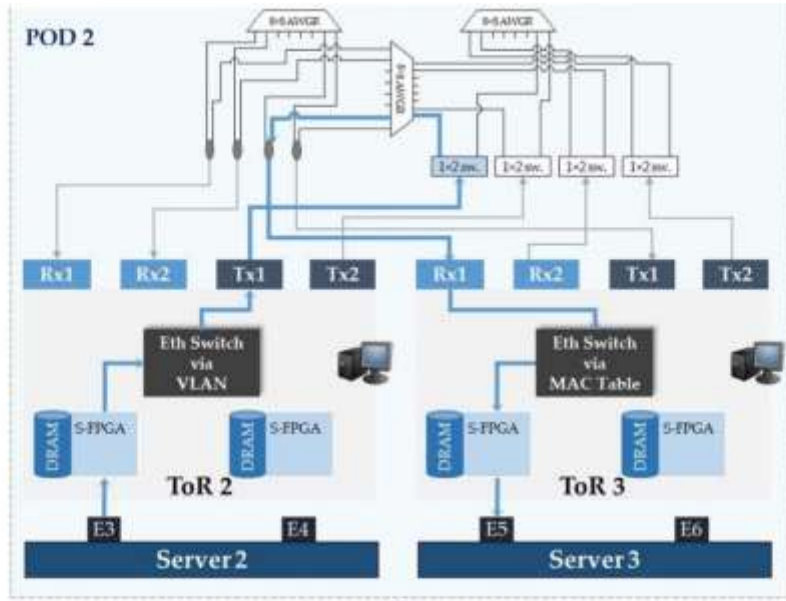
The second scenario implemented in the demo was focused on the *intra-Pod* communication. In this scenario, the two ToRs, which reside within the POD 2, are establishing a communication path. The generated traffic originates from E3 host of ToR 2 and its destination is E5 host that belongs to ToR 3. The experiment was carried out with three different traffic patterns; 20, 40 and 70 slots of the total 80 of the Scheduling Period. In this way, the bandwidth is consecutively increased. The commands for increasing the allocated slots are sent by the SDN controller to the FPGAs that enforce the instructions in real-time.

This experiment involves three data plane modules: two ToR switches and one POD switch. On Figure 11 (a) a schematic diagram of the traffic route through the data plane is presented. Following the blue arrows, the frames originating from Server E3 are first buffered in the S-FPGA of ToR 2. After that, the traffic is routed through the Mellanox Ethernet Switch, which selects the right N-FPGA as the destination gateway to the optical rings of the proposed network.

The tunable transmitter driven by the N-FPGA of ToR 2 is tuned to transmit the traffic enrolled in the correct wavelength according to the destination ToR; in this case  $\lambda_3=1548.515\text{nm}$  (Table 1). The transmitter is followed by a  $1\times 2$  optical switch which is responsible to handle the frames in accordance to their destination place (intra-pod or inter-pod). In this case the  $1\times 2$  optical switch, which is driven according to the schedule by the N-FPGA, routes the frames to the output that routes to the same POD. The cyclic  $8\times 8$  AWG forwards passively the frames to the fiber that reaches the receiver embedded to the N-FPGA of ToR3. It is important to mention that due to the system optimization in the optical path (and in all intra-pod optical paths) there is no need for optical amplification.

After the successful optical reception of the frames by the N-FPGA of ToR 3, the extracted Ethernet frames are forwarded to the Mellanox switch, which in turn routes the Ethernet Frames to the correct S-FPGA using the destination's preregistered MAC address. The S-FPGA forwards, without any additional delay or processing, the Ethernet Frames to the destination E5.

The "monitor" images at the bottom of Figure 11 (b-d) show screenshots of the frames traveling through the network taken by means of a real-time oscilloscope. In these figures, the screenshots represent different percentages of slot allocation. It is useful to repeat that the NEPHELE scheduling period is divided into 80 slots (200 us each slot) with a guard time (10 us) between them. Apparently, the increase of slots that are allocated for a specific connection results to a direct increase of the available bandwidth between the two endpoints. Figure 11 (b) and (c), (d) present the *nload* command running on all the involved Servers. The bandwidth scales according to the slot allocation.



**Figure 11:** (a) The data-plane structure of the intra-pod scenario. The frames follow the route highlighted with the blue arrows. (b-d) Screenshots of the frames sent from ToR2 Server E3 (10.2.2.1) and received by ToR3 server E5 (10.2.3.1): (b) In this case 25% of the scheduling period (20 slots out of 80 slots) is occupied for the communication path between ToR2 and ToR3 achieving 1.8Gbit/s bandwidth. (c) In this half of the scheduling period (40 slots out of 80 slots) is occupied for the communication path between ToR2 and ToR3 achieving 3.6 Gb/s bandwidth. (d) In this case 87.5% of the scheduling period (70 slots out of 80 slots) is occupied for the communication path between ToR2 and ToR3 achieving 6.4 Gbit/s bandwidth.

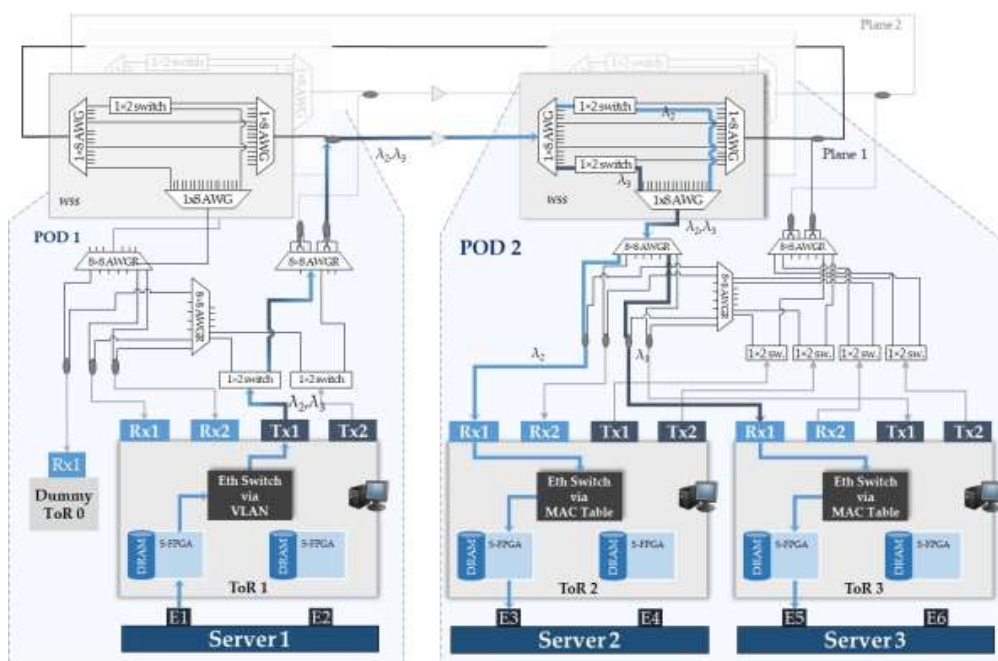
#### 4.1.3 Inter-Pod scenario I

The experimental setup used for the inter-pod communication is shown in Figure 12 (for this section, we assume only the light-blue arrows in the figure, while the dark-blue ones correspond to the scenario in section 4.1.4). In this case server E1 which resides inside ToR 1 of POD 1 establishes connection with server E3 which is connected to ToR 2 of POD 2. Thus, in this scenario two POD switches and two ToR switches are involved. Also, the

implemented traffic pattern includes 60 out of the 80 slots, which is the total Scheduling Period.

The tunable transmitter is tuned to emit the wavelength which will reach ToR 2; i.e. the wavelength  $\lambda_2= 1547.715\text{nm}$ . The transmitted traffic passes through all the optical components of two consequent POD switches: 1x2 switch (intra-, inter-), 8x8 AWGR, EDFA, WSS, 8x8 AWGR and a coupler.

Figure 13 depicts the packets screenshots taken by means of a real-time oscilloscope and a constant slot allocation (60 out of 80 slots). In addition, it presents the *nload* command running on the Servers and shows that the achieved bandwidth between E1 and E3 is 5.12Gbit/s.



**Figure 12:** Experimental setup for the inter-pod scenario; dark and light blue lines reveal the optical paths from source host (E1) to destination hosts (E3 and E5) respectively.



**Figure 13:** (a) The packets sent from E1 (10.1.1.1) ToR1 and received by E3 (10.2.2.1)ToR2. In this case 75% of the scheduling period (60 slots out of 80 slots) is



occupied for the communication path between ToR1 and ToR2 achieving 5.12 Gb/s bandwidth.

#### 4.1.4 Inter-Pod scenario II

In this scenario, the traffic follows an inter-pod route as well. In this case though, the server of ToR1, which resides inside POD 1, sends traffic to two different ToRs in POD 2. More specifically, server E1 (ToR 1 – POD 1) generates and sends traffic to server E3 (ToR 2 – POD 2) and E5 (ToR 3 – POD 2) alternately. The source server (E1) is transmitting Ethernet Frames for both destinations concurrently while the scheduling commands, sent by the SDN controller, impose the slot allocation for the scheduling period. The experimental setup for the inter-pod scenario with two destination ToRs is depicted in Figure 12. The traffic originating from E1 interface is optically transmitted by Tx1 of ToR 1 and it is forwarded into the optical ring of plane 1 by means of a 1×2 optical switch (intra- or inter-pod) and an 8x8 AWGR. As mentioned above, the tunable transmitter Tx1 is creating two traffic flows enrolled in two wavelengths alternatingly according to the targeted ToR;  $\lambda_2=1547.715$  nm for ToR 2 and  $\lambda_3=1548.515$  nm for ToR 3. An optical amplifier (EDFA) is used for compensating the optical losses between the two PODs.

The Wavelength Selective Switch (WSS) of POD 2, which is the responsible module for dropping specific wavelengths from the WDM rings, forwards the optical flows towards its ToRs according to the SDN controller’s commands. The switches inside the WSS are driven by the N-FPGA with TTL signals. An additional passive AWGR, as shown in Figure 12, finally routes the traffic to ToR 2 and ToR 3.

The green trace depicted in Figure 14 (d) represents the frames that reach ToR 2 and the blue one the corresponding frames which are received by ToR 3. It is obvious that the NEPHELE traffic is divided into two equal parts for each destination and more specifically, 25% of the scheduling period (20 slots out of 80) is occupied for each of the two ToRs (Figure 14 (a-c)). Finally, the bandwidth reached almost 1.4 Gb/s at both servers E3 and E5, as it is shown in Figure 14.



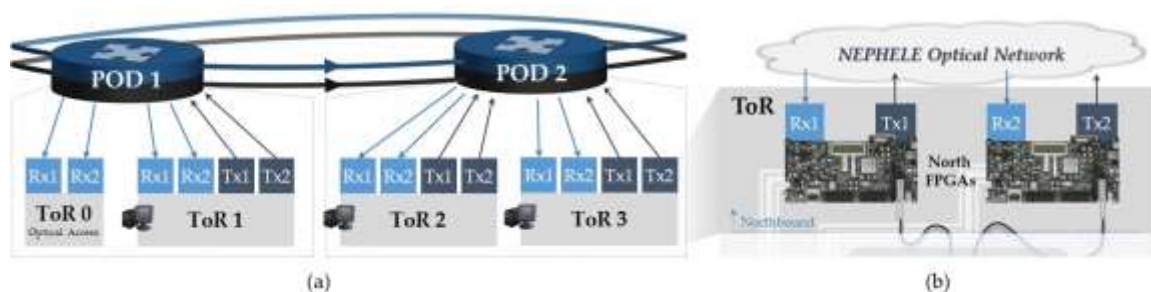
**Figure 14:** (a) Server E1 (10.1.1.1) is sending data traffic to (b) server E3 (10.2.2.1) and (c) server E5 (10.2.2.129) achieving almost 1.4 Gb/s bandwidth at each server. (d) Screenshots of the NEPHELE Frames originating from ToR1 and captured in ToR 2 (blue trace) and

ToR 3 (green trace). The total traffic is divided equally into two destination ToRs (20 slots reaching each ToR).

#### 4.1.5 Combined Intra-pod and Inter-pod communication scenarios

The following sections present additional, more complex scenarios that were tested on the developed innovative mini-datacenter testbed. Extra functionalities were implemented in order to confirm the systems' stability despite the additional complexity.

To begin with, the scenarios that are presented in the next sections were carried out on the same demo testbed (Figure 15 (a)) with a main difference from the previous ones: communication between the ToRs takes place from the northbound interfaces of each ToR switch as it is depicted in Figure 15 (b). This means that the modules located vertically below the north FPGAs are not involved in these experiments.



**Figure 15:** (a) The experimental setup used for the evaluation of the combined Intra-pod and Inter-pod Scenarios, (b) The northbound data plane interfaces schematic of the ToR switches.

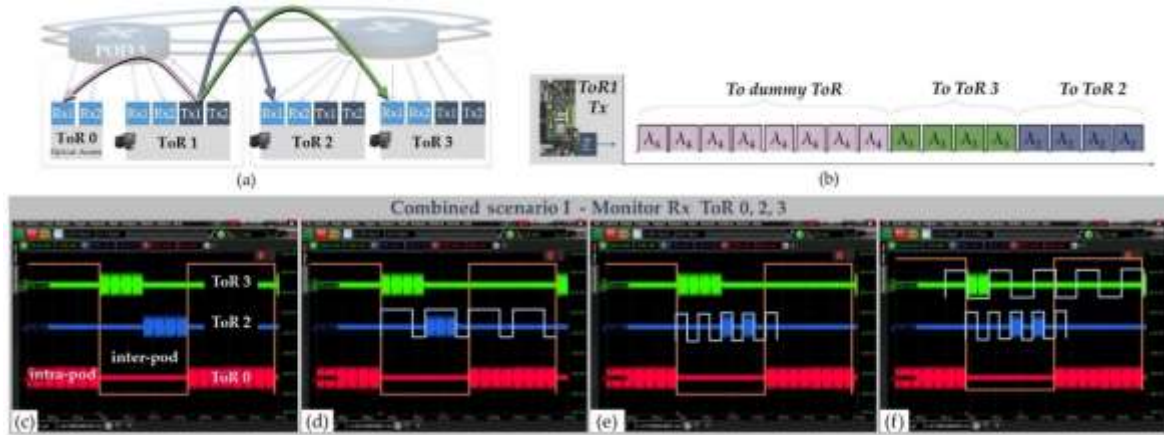
More specifically, the N- FPGA generates “dummy” frames filled with PRBS instead of Ethernet data. This dummy traffic meets all the specifications of the proposed architecture with respect to timing and rate. Also, the north FPGA generates and provides the control signals (TTL) for the optical modules which are involved: 1×2 optical switches for the intra- and inter-pod communication and 1×2 WSS (the “Demultiplex, switch and multiplex” approach of the WSS module includes multiple 1×2 optical switches).

##### 4.1.5.1 Combined Scenario I

The first scenario which combines intra-pod and inter-pod traffic is shown in Figure 16 (a). According to this, ToR 1 of POD 1 sends traffic to the ToR 0 of POD 1 (intra-pod communication) and to ToR 2 and ToR 3 of POD 2 (inter-pod communication). Hence, the tunable transmitter of ToR 1 is set to transmit sixteen repeated frames enrolled in different wavelengths (in  $\lambda_4$ ,  $\lambda_2$  and  $\lambda_3$ ) as it is shown in Figure 16 (b).

Figure 16 (c-f) portrays the screenshots captured after the receivers by means of a real-time oscilloscope. The oscilloscope's channels are connected to each of the three photoreceivers; the red trace corresponds to ToR 0, the blue one to ToR 2 and the green one to ToR 3. In all

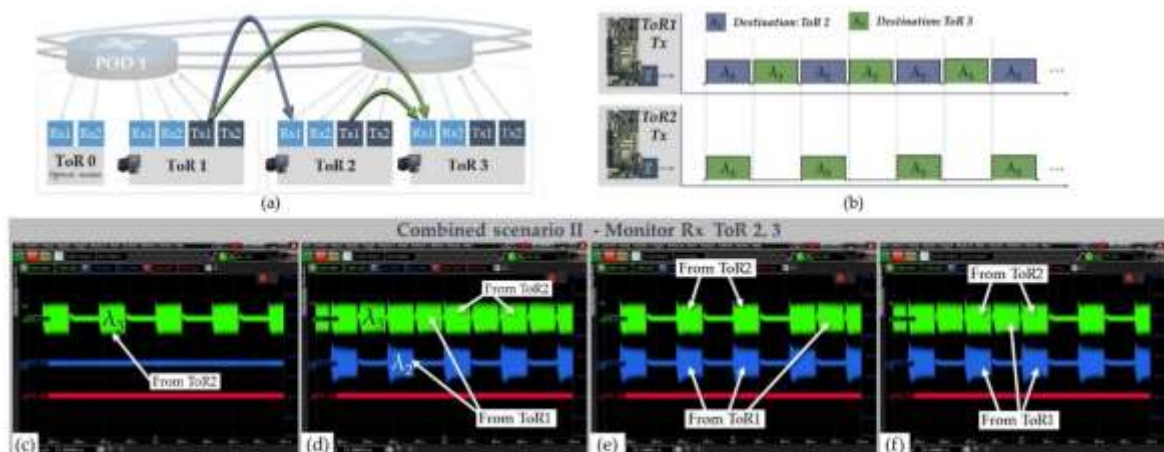
the screenshots (c-f) the orange trace represents the driving signal that is applied to the 1×2 optical switch which distinguishes the intra- and inter-pod traffic in POD 1. In addition, the white traces in (d-f) screenshots represent the driving signal for the 1×2 optical switches which are located inside the WSS of POD 2 and correspond to  $\lambda_2$  and  $\lambda_3$ .



**Figure 16:** (a) Combined scenario I. (b) The repeating sequence of frames in three wavelengths produced by the tunable Tx of ToR 1: eight frames in  $\lambda_4$ , four in  $\lambda_2$  and four in  $\lambda_3$ . (c-f) Screenshots of the received frames to each ToR receiver; red trace corresponds to the ToR 0 receiver, blue trace to the ToR 2 receiver and green trace to the ToR 3 receiver.

#### 4.1.5.2 Combined Scenario II

The second combined scenario that was implemented is presented in Figure 17 (a). As shown, ToR 1 of POD 1 transmits frames to ToR 2 and ToR 3 of POD 2 (inter-pod communication). At the same time and synchronously, the transmitter of ToR 2 sends traffic to ToR 3 (intra-pod communication). As a result, the receiver of ToR 3 receives traffic from both ToR 1 and ToR 2.



**Figure 17:** (a) Schematic for the Combined Intra-pod and Inter-pod Scenario II, (b) Repeating sequence of frames transmitted by ToR 1 and ToR 2 simultaneously and

synchronously at  $\lambda_2$  and  $\lambda_3$ , (c-e) Screenshots of the received frames to ToR 2 and ToR 3 receivers; blue and green traces represent the receivers at ToR 2 and ToR 3 respectively.

The traffic generated by ToR 1 (POD 1) and ToR 2 (POD 2) is depicted in Figure 17 (b). The tunable transmitter of ToR 1 is able to produce a sequence of frames at  $\lambda_2$  and  $\lambda_3$ . Similarly, ToR 2 transmitter is programmed to generate frames at  $\lambda_3$  with an empty slot between them in order to avoid any possible conflict with  $\lambda_3$  frames transmitted by ToR1.

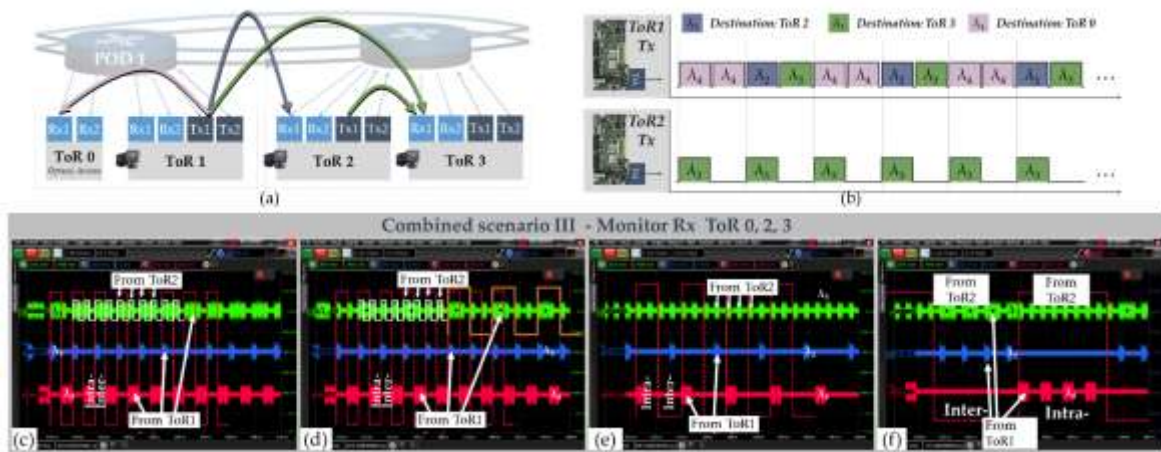
Combinations of the above-mentioned traffic flows are presented in the screenshots (c-f) of Figure 17. Firstly, Figure 17 (c) shows the frames at  $\lambda_3$  that reach ToR 3 and are transmitted by ToR 2, without any other transmitted traffic. Afterwards, the transmitter of ToR 1 starts sending frames at  $\lambda_2$  and  $\lambda_3$  as observed in Figure 17 (d). In this case, the  $1 \times 2$  switch of POD 1 forwards the traffic to the inter-pod path and the WSS of POD 2 drops the two wavelengths. The next two screenshots (Figure 17 e-f) show cases in which the WSS of POD 2 drops some of the frames at  $\lambda_3$ .

#### 4.1.5.3 Combined Scenario III

The third and final scenario that was carried out on the demo testbed is a combination of the two previous scenarios and it is shown in the schematic of Figure 18 (a). The traffic produced by ToR1 is both intra-pod (ToR 0 - POD 1) and inter-pod (ToR 2 and ToR 3 – POD 2). In the meantime, ToR 2 generates another intra-pod traffic flow, but this time inside POD 2 by transmitting frames with destination the ToR 3. In addition, three wavelengths are involved ( $\lambda_2$ ,  $\lambda_3$  and  $\lambda_4$ ) in this scenario and obviously the two transmitters are synchronised.

The traffic that is generated by the two tunable transmitters is depicted in Figure 18 (b). The transmitter of ToR 1 (POD 1) creates a sequence of frames enrolled in  $\lambda_4$ ,  $\lambda_2$ ,  $\lambda_3$ . As it happened in the previous combined scenario, the ToR 2 (POD 2) transmitter is programmed to generate frames in  $\lambda_3$  with an empty slot between them.

Figure 18 (c-f) represent the screenshots of the frames that reach the three destination ToRs. The red frames represent the ones at  $\lambda_4$  that are received by the ToR 0 Rx. Similarly, the blue and green traces show the frames that arrive to ToR 2 and ToR 3 respectively. It is important to mention that in all screenshots the two transmitters are set to emit the sequences shown in Figure 18 (b). In addition, the dotted red waveform indicates the separation of intra- and inter-pod traffic of POD 1. White arrows and transmitter's labels are highlighted on the following figures indicating the origin of each packet. More specifically, in Figure 18 (c) and (d) the frames transmitted by ToR 2 (intra-pod inside POD 2) are shown in the green waveform. Also, the orange trace and the black bullets highlighted in Figure 18 represent the WSS on/off state for  $\lambda_3$  and the frames transmitted by ToR 1 respectively.

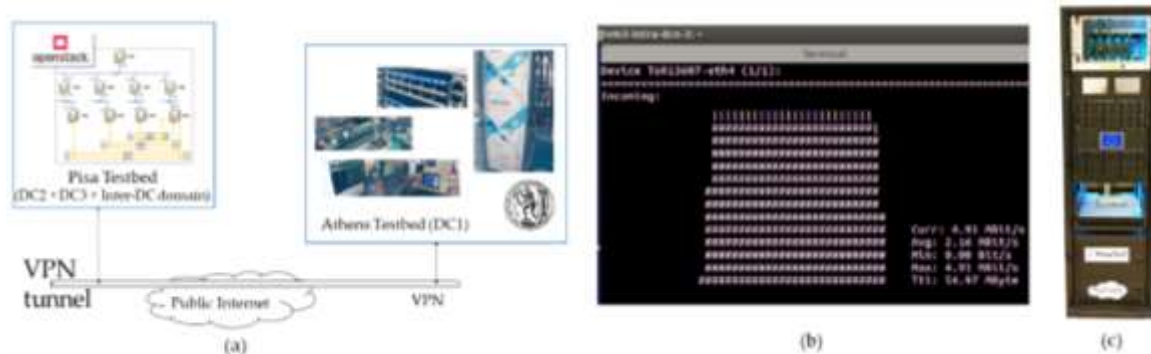


**Figure 18:** (a) Schematic for the Combined Intra-pod and Inter-pod Scenario III, (c-e) Screenshots of the received NEPHELE frames to each ToR receiver; red trace corresponds to the dummy ToR receiver, blue trace to the ToR 2 receiver and green trace to the ToR 3 receiver

#### 4.2 Inter-Datacenter Communication - Pisa and Athens Testbed communication

In order to verify the capability of the framework to control heterogeneous technologies, we validated it in a multi-site inter-DC testbed deployed at IRT Testbed and NTUA facilities, in Pisa Italy and Athens Greece, respectively. The environment includes two DCNs emulated by Mininet, connected through an inter DC network (also emulated) to a third physical DCN, based on the data plane. The emulated networks are running in Virtual Machines (VMs) placed in IRT testbed, while the physical proposed DCN is located at NTUA premises (see Figure 19 (a)). The two sites are interconnected through a VPN tunnel instantiated between two hosts acting as gateways, with the VPN playing the role of an inter-domain link.

This environment combines physical and emulated network domains and allows us to verify the correct cooperation between control and data plane for both intra-DC and inter-DC scenarios. Moreover, we are also able to demonstrate the applicability of the system to scenarios involving multiple network domains deploying different technologies, a key feature for the backward compatibility of solution and its possible deployment in existing multi-DC infrastructures.



**Figure 19:** (a) Multi-site testbed for integrated intra-DC and inter-DC tests. (b) The bandwidth transported over the public internet was measured at 4.91 Mbit/s. (c) The industrial demonstrator of complementing the Athens Testbed demonstrator including the presented technologies

The experiment is triggered through the NIDO GUI sending a request for an end-to-end path between an emulated server located in the IRT testbed and a physical server in the NTUA testbed. At the control plane level, the interaction between NIDO and the SDN controllers OCEANIA and JULIUS, as well as the exchange of OpenFlow messages between the controllers and the data plane, are executed and the end-to-end connection is successfully established.

In order to verify the actual connectivity between the two servers, we use the *iPerf* tool to generate traffic between the servers and we verify the exchanged traffic using the *tcpdump* tool. Although the bandwidth reported (4.91Mbit/s - Figure 19 (b)) is very low for NEPHELE standards, the traffic is transported through the VPN over the public Internet, so the actual transfer rate is not a significant indication of the performance of the physical DCN infrastructure.

Finally, the industrial demonstrator of complementing the demonstrator in Athens and the proposed technologies is depicted in Figure 19 (c).

- 
- i. Cisco, "Cisco Annual Internet Report (2018–2023)", white paper, Available online: <https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/white-paper-c11-741490.html> (accessed on 20/03/2020)
  - ii. A. Singh, et. al. "Jupiter Rising: A Decade of Clos Topologies and Centralized Control in Google's Datacenter Network", SIGCOMM 2015.
  - iii. W. M. Mellette, "A Practical Approach to Optical Switching in Data Centers," 2019 Optical Fiber Communications Conference and Exhibition (OFC), San Diego, CA, USA, 2019, pp. 1-3.

- 
- iv. D. Zang, M. Chen, N. Sun, R. Proietti, S. J. B. Yoo and Z. Cao, "OpticV: An energy-efficient datacenter network architecture by MEMS-based all-optical bypassing," 2016 IEEE Optical Interconnects Conference (OI), San Diego, CA, 2016, pp. 70-71.
  - v. Y. B. Ezra and B. I. Lembrikov, "Application of All-Optical Memory for Advanced Modulation Formats in Communication Intra-Datacenter Networks (Intra-DCNs)," 2018 20th International Conference on Transparent Optical Networks (ICTON), Bucharest, 2018, pp. 1-4.
  - vi. Z. Zhong, C. Guo and G. Shen, "Scheduling Traffic Switching in An All-Optical Intra-Datacenter Network with Sub-Waveband Switching," 2018 Asia Communications and Photonics Conference (ACP), Hangzhou, 2018, pp. 1-3.
  - vii. K. Sato et al., "A Large-Scale Wavelength Routing Optical Switch for Data Center Networks," IEEE Commun. Mag., vol. 51 no. 9, Sept. 2013, pp. 46-52.
  - viii. G. M. Saridis et al., "Lightness: A Function-Virtualizable Software Defined Data Center Network With All-Optical Circuit/Packet Switching," in Journal of Lightwave Technology, vol. 34, no. 7, pp. 1618-1627, 1 April, 2016.
  - ix. X. Xue et al., "SDN-Enabled Reconfigurable Optical Data Center Networks Based on Nanoseconds WDM Photonics Integrated Switches," 2019 21st International Conference on Transparent Optical Networks (ICTON), Angers, France, 2019, pp. 1-4.
  - x. Shuping Peng, Bingli Guo, Chris Jackson, Reza Nejabati, Fernando Agraz, Salvatore Spadaro, Giacomo Bernini, Nicola Ciulli, and Dimitra Simeonidou, "Multi-Tenant Software-Defined Hybrid Optical Switched Data Centre," J. Lightwave Technol. 33, 3224-3233 (2015)
  - xi. P. Bakopoulos et al., "NEPHELE: An End-to-End Scalable and Dynamically Reconfigurable Optical Architecture for Application-Aware SDN Cloud Data Centers," in IEEE Communications Magazine, vol. 56, no. 2, pp. 178-188, Feb. 2018.
  - xii. P. Bakopoulos, K. Tokas, C. Spatharakis and H. Avramopoulos, "Slotted optical datacenter network with sub-wavelength resource allocation," 2017 IEEE Photonics Society Summer Topical Meeting Series (SUM), San Juan, 2017, pp. 161-162.
  - xiii. K. Tokas, I. Patronas, C. Spatharakis, D. Reisis, P. Bakopoulos and H. Avramopoulos, "Slotted TDMA and optically switched network for disaggregated datacenters," 2017 19th International Conference on Transparent Optical Networks (ICTON), Girona, 2017, pp. 1-5.
  - xiv. G. Landi et al., "SDN control framework with dynamic resource assignment for slotted optical datacenter networks," 2017 Optical Fiber Communications Conference and Exhibition (OFC), Los Angeles, CA, 2017, pp. 1-2.
  - xv. K. Christodoulopoulos, K. Kontodimas, A. Siokis, K. Yiannopoulos and E. Varvarigos, "Efficient bandwidth allocation in the NEPHELE optical/electrical datacenter interconnect," in IEEE/OSA Journal of Optical Communications and Networking, vol. 9, no. 12, pp. 1145-1160, Dec. 2017.
  - xvi. K. Tokas et al., "Real Time Demonstration of an End-to-End Optical Datacenter Network with Dynamic Bandwidth Allocation," 2018 European Conference on Optical Communication (ECOC), Rome, 2018, pp. 1-3.
  - xvii. C. Spatharakis, K. Tokas, I. Patronas, P. Bakopoulos, D. Reisis and H. Avramopoulos, "NEPHELE: Vertical Integration and Real-Time Demonstration of an Optical Datacenter Network," 2018 20th International Conference on Transparent Optical Networks (ICTON), Bucharest, 2018, pp. 1-4.
  - xviii. NIDO orchestrator. Available online: <https://github.com/nextworks-it/nephele-nido> (accessed 25/03/2020)
  - xix. G. Landi, M. Capitani, A. Kretsis, P. Kokkinos, K. Christodoulopoulos and E. Varvarigos, "Joint Intra- and Inter-Datacenter Network Optimization and Orchestration," 2018 Optical Fiber Communications Conference and Exposition (OFC), San Diego, CA, 2018, pp. 1-3.
  - xx. OCEANIA SDN controller. Available online: <https://github.com/nextworks-it/oceania-dcn-controller> (accessed 25/03/2020)
  - xxi. A. Kretsis, L. Corazza, K. Christodoulopoulos, P. Kokkinos and E. Varvarigos, "An emulation environment for SDN enabled flexible IP/optical networks," 2016 18th International Conference on Transparent Optical Networks (ICTON), Trento, 2016, pp. 1-4.

- 
- xxii. I. Patronas, N. Gkatzios, V. Kitsakis, D. Reisis, K. Christodouloupoulos and E. Varvarigos, "Scheduler Accelerator for TDMA Data Centers," 2018 26th Euromicro International Conference on Parallel, Distributed and Network-based Processing (PDP), Cambridge, 2018, pp. 162-169.
- xxiii. Photonics Communications Research Laboratory website. Available Online: <https://photonics.ntua.gr> (accessed 05/04/2020)
- xxiv. QLogic 57810 Dual Port 10Gb Network Adapter, Product Specifications. Available online: <https://www.dell.com/en-my/shop/qlogic-57810-dual-port-10gb-direct-attach-sfp-network-adapter-full-height/apd/540-bbgs/networking> (accessed: 25/03/2020)