

High-Speed Optical Systems for Intra- and Inter-Datacenter Networking



Mohammed Sowailem

Department of Electrical & Computer Engineering
McGill University
Montréal, Canada

February 2017

A thesis submitted to McGill University in partial fulfillment of
the requirements of the degree of Doctor of Philosophy

© 2017 Mohammed Sowailem

Abstract

The ever-growing capacity demand for datacenter interconnects (DCIs), for intra- and inter-datacenter networking, controls the development of the cutting-edge software/hardware technologies in data communications. However, the concern of these applications is not only the capacity, but also the cost, power consumption, complexity, and the form factor of the deployed transceivers. This makes investing in integrated optoelectronic devices important. Specifically, InP-based platform is a promising state-of-the-art technology to achieve high capacity transmission and build small form factor transceivers. Another challenge for intra-datacenter networking is to provide a full bisection bandwidth in core layer switching. Providing full bisection bandwidth using electronic switching is challenging due to limited bandwidth, wiring complexity, and power consumption. Accordingly, the use of optical switching is necessary not only to meet current challenges, but also to provide reliable solution for next generation intra-datacenter networking.

In this thesis, we aim at investigating and proposing possible optical solutions for DCIs operating in intra- and inter-datacenter networking. In intra-datacenter networking, we propose possible network architectures for practical implementation of optical core switching. More specifically, we address the system level and associated control plane requirements, and how the physical layer optical components operate within this proposed network architecture. Next, we move deeper into the physical layer and propose an optical solution that increases the capacity per wavelength taking into consideration system complexity. On the other hand for inter-datacenter networking, we demonstrate the viability of deploying InP-based dual polarization in-phase/quadrature (IQ) modulator (InP-based DP-IQM) in optical transceivers of high speed DCIs. The abovementioned DCI solution is a high-speed optical coherent transmission system enabled by the use of two successive generations of digital-to-analog converter (DAC) technology and digital signal processing (DSP) that handle transceiver impairments. In addition, we investigate a two-dimensional Stokes vector direct detection (2D-SV-DD) system as an alternate solution for inter-datacenter applications with reduced complexity at the expense of halving the capacity compared to coherent systems.

In the first part of this thesis, we focus on intra-datacenter networking including how optical switching can be implemented in terms of network architecture and switching paradigm. We propose an arrayed waveguide grating router based (AWGR-based) switching mechanism for passive optical switching in an optical burst switching (OBS) datacenter. In addition, we propose a contention resolution strategy that uses electronic assembly buffers at the transmitter for both burst construction and contention resolution. Next, we propose a self-homodyne system for intra-datacenter interconnects using small form factor InP-based DP-IQM. The main idea of the proposed system is to benefit from the full duplex fibers connecting the optical transceivers within the datacenter to enable coherent reception with reduced complexity. We report data rate of 448 Gb/s and 320 Gb/s on a single wavelength below KP4 forward error correcting threshold at 2 km and 10 km, respectively.

In the last part of the thesis, we focus on inter-datacenter interconnects deploying InP-based DP-IQM in optical coherent transmission and 2D-SV-DD system. We build a test bed for optical coherent transmission for InP-based DP-IQM and evaluate the viability of using this integrated technology for 400G and beyond next generation systems for DCIs. We demonstrate experimentally, using an InP-based DP-IQM on a single wavelength, 448 Gb/s (56 Gbaud PDM-16QAM) and 770 Gb/s (77 Gbaud PDM-32QAM) transmission over 500 km and 320 km enabled by DACs operating at sampling rates of 65.7 GSps and 84 GSps, respectively. Next, we explore the use of the same InP device in the less complex 2D-SV-DD system. We demonstrate experimentally the transmission of single carrier 56 Gbaud 16QAM, 8QAM, and QPSK optically modulated signals over 320, 960, and 2,880 km, respectively, enabled by 64 GSps DAC and using chromatic dispersion (CD) pre-compensation at the transmitter side. In addition, we experimentally demonstrate using 84 GSps DAC, aided with mathematical analysis, two suitable methods for CD compensation, namely CD pre- and post-compensation for 2D-SV-DD system. Our findings reveal that CD pre-compensation can be used for any transmission distance, where we are able to report the highest throughput-times-distance product achieved in Stokes transmission systems. For example, 84 Gbaud QPSK and 64 Gbaud QPSK transmissions reached 3,520 km and 4,800 km below the bit error rate (BER) of 2×10^{-2} achieving 591,360 Gb/s.km and 614,400 Gb/s.km throughput-times-distance products, respectively.

Résumé

La demande croissante de capacité pour les interconnexions dans les centres de données (ICD), pour les réseaux intra- et inter-centres de données, contrôle le développement des technologies logicielles / matérielles de pointe pour la transmission de données. Cependant, le focus de ces applications n'est pas seulement la capacité, mais aussi le coût, la consommation d'énergie, la complexité et le facteur de forme des émetteurs-récepteurs déployés. Il est donc important d'investir dans les dispositifs optoélectroniques intégrés. Plus précisément, la plate-forme basée sur InP est une technologie de pointe prometteuse pour la transmission à large capacité et la construction d'émetteurs-récepteurs à facteur de forme de petite taille. Un autre défi pour les réseaux intra-centres de données est de fournir une largeur de bande de bisection complète pour la commutation de la couche centrale. Fournir une largeur de bande de bisection complète en utilisant la commutation électronique est difficile en raison de la bande passante limitée, la complexité du câblage et la consommation d'énergie. En conséquence, l'utilisation de la commutation optique est nécessaire non seulement pour relever les défis actuels, mais aussi pour fournir une solution fiable pour la prochaine génération de réseaux intra-centres de données.

Dans cette thèse, nous visons à étudier et à proposer des solutions optiques possibles pour les ICD intra- et inter- centres de données. Pour les réseaux intra-centres de données, nous proposons des architectures de réseau possibles pour la mise en application pratique de la commutation optique de la couche centrale. Plus précisément, nous abordons le niveau système et les exigences associées pour le plan de commande, et comment les composants optiques de la couche physique fonctionnent sous cette architecture de réseau proposée. Ensuite, nous allons plus profond dans la couche physique et proposons une solution optique qui augmente la capacité par longueur d'onde en tenant compte de la complexité du système. D'autre part, pour les réseaux inter-centres de données, nous démontrons la viabilité du déploiement du modulateur en phase / quadrature (IQ) à double polarisation basé sur InP (MIQ-DP sur plateforme InP) dans les émetteurs-récepteurs optiques des ICD à grande vitesse. La solution ICD susmentionnée est un système de transmission optique cohérent à grande vitesse, rendu possible grâce à l'utilisation de deux générations successives de convertisseurs digital-analogique (CDA) et du traitement numérique des signaux (DSP) qui gèrent les dégradations de l'émetteur-récepteur. De plus, nous

étudions un système bidimensionnel de détection directe de vecteurs de Stokes (2D-DD-VS) comme solution alternative pour les applications inter-centres de données, avec une complexité réduite mais au détriment d'une réduction de la capacité par rapport aux systèmes cohérents.

Dans la première partie de cette thèse, nous nous concentrons sur les réseaux intra-centres de données, notamment sur la façon dont la commutation optique peut être mise en œuvre en termes d'architecture de réseau et de paradigme de commutation. Nous proposons un mécanisme de commutation basé sur un routeur réseau à guides d'ondes (AWGR) pour la commutation optique passive dans un centre de données avec commutation en rafale (OBS). De plus, nous proposons une stratégie de résolution de contention qui utilise des tampons d'assemblage électroniques à l'émetteur pour la construction de rafales et la résolution de contention. Ensuite, nous proposons un système auto-homodyne pour les interconnexions intra-centres de données, utilisant un MIQ-DP à petit facteur de forme basé sur InP. L'idée principale du système proposé est de bénéficier des fibres bidirectionnelles simultanées reliant les émetteurs-récepteurs optiques au sein du centre de données pour permettre une réception cohérente avec une complexité réduite. Nous rapportons un débit de données de 448 Gb / s et 320 Gb / s sur une seule longueur d'onde sous le seuil de correction d'erreur directe KP4 à 2 km et 10 km, respectivement.

Dans la dernière partie de la thèse, nous nous intéressons aux interconnexions inter-centres de données qui déploient le MIQ-DP basé sur InP dans un système de transmission optique cohérente avec 2D-DD-VS. Nous construisons un banc d'essai pour la transmission cohérente optique avec le MIQ-DP basé sur InP et évaluons la viabilité de l'utilisation de cette technologie intégrée pour les systèmes de prochaine génération 400G et au-delà pour les ICD. Nous démontrons expérimentalement, en utilisant un MIQ-DP basé sur InP à une seule longueur d'onde, une transmission de 448 Gb / s (56 Gbaud PDM-16QAM) et 770 Gb / s (77 Gbaud PDM-32QAM) sur 500 km et 320 km rendue possible par des CDA fonctionnant à des taux d'échantillonnage de 65,7 GSps et 84 GSps, respectivement. Ensuite, nous explorons l'utilisation du même MIQ-DP basé sur InP dans le système 2D-DD-VS, moins complexe. Nous démontrons expérimentalement la transmission de signaux optiques modulés à 56 Gbaud sur une seule longueur d'onde aux formats 16QAM, 8QAM et QPSK sur 320, 960 et 2 880 km respectivement, avec un CDA à 64 GSps et en utilisant la pré-compensation de dispersion chromatique (DC) au niveau de l'émetteur. De plus, nous démontrons expérimentalement, en utilisant un CDA à 84 GSps et une analyse mathématique, deux méthodes appropriées pour la compensation de DC, à

savoir la pré et post-compensation DC pour le système 2D-DD-VS. Nos résultats révèlent que la pré-compensation de DC peut être utilisée pour toute distance de transmission, où nous sommes en mesure de rapporter le produit débit-distance le plus élevé atteint dans les systèmes de transmission Stokes. Par exemple, des transmissions de 84 Gbaud QPSK et 64 Gbaud QPSK ont atteint 3 522 km et 4 800 km en-dessous du taux d'erreur binaire (TEB) de 2×10^{-2} , atteignant respectivement des produits débit-distance de 591 360 Gb / s.km et de 614 400 Gb / s.km.

Acknowledgments

Praise is to ALLAH the Most Merciful the Most Generous. As Prophet Muhammad peace be upon him taught us that whoever is not thankful to people, he indeed is not thankful to ALLAH.

First, I would like to thank my supervisors; Prof. David V. Plant and Prof. Odile Liboiron-Ladouceur for their academic advice, constant encouragement, guidance, patience, and support since starting my Ph.D. I want to express my deepest gratitude to Prof. David V. Plant for his continuous and endless support during the ups and downs of the path of my research and my personal life. I greatly appreciate his tremendous efforts that keep the laboratories equipped by cutting-edge technologies and state-of-the-art components allowing us to explore our capabilities of having impact not only in the academic community, but in industry as well. He has the ability to motivate us for work through humbleness and kindness rather than pressure and stress. He was always at my side when I need his support as a supervisor and in my personal life. Thanks a lot for everything; being your student is a privilege. Also, I would like to thank Prof. Odile Liboiron-Ladouceur for her patience and support especially during the first phase of my research. Unlike any other co-supervisor, she gave exceptional attention to my work improving technical representation of my publications through fruitful discussions and feedback. Also, I greatly appreciate her support and efforts to accomplish the thesis in its last shape.

I would like to express my gratitude to many talented individuals who helped me throughout the course of my Ph.D. Many thanks to my friend and lab mate Thang Hoang with whom I spent countless hours in the lab over the years. His contribution is spread all over this thesis since I did most of my experimental work with him. We did great work as a team that I would have never accomplished individually. Many thanks to my friend or actually my brother Dr. Mohamed Osman for transferring his lab experience to us, his technical support, his fruitful and hours of discussions and paving my way to start the second phase of my research. Thanks to Dr. Mathieu Chagnon for his support and especially for his time spent on response to reviewers' comments. Thanks to my colleagues and friends in our group; Eslam El-fiky, Meng Qiu, David Patel, and Dr. Mahdi Malekiha for their support in my work. Thanks to Maxime Jacques for his help translating the abstract into French. Thanks to all our predecessors who initially built the basic test bed in our lab and wrote the basic code for our experiments; Dr. Xian Xu, Dr. Mohamed

Osman, Dr. Qunbi Zhuge, Dr. Benoît Châtelain, Dr. Ziad El-Sahn, Dr. Mathieu Chagnon and Dr. Mohammad Pasandi. Thanks to all PSGers for their help whenever needed. I am indebt to current and former administrative members of Sytacom for their help in relieving us from a lot of overhead related to purchasing components, shipment, paper works and IT services.

Many thanks to TeraXion, Inc. who started the collaboration on the InP device project and to Ciena, Corp. for continuing the project after the department acquisition. Special thanks to their team Mr. Carl Paquet, Mr. Stéphane Paquet, Mr. Michel Cyr and Dr. Ian Woods. Thanks continue to Ciena, Corp. members Dr. Charles Laperle and Dr. Qunbi Zhuge for their fruitful feedbacks and unpaid inputs regarding my publications.

I would like to express my deep appreciation and pay humble respects to my parents Eng. Salwa Fadel and Eng. Yousry Sowaillem. Any achievement in my life is just a minor part of your achievements starting from my existence in this life and continuous guidance, love, support, and prayers. As Prophet Muhammad peace be upon him mentioned what can be translated to mean that we will never fulfill your rights upon us. I am deeply sorry that you are not here with me to share with me those moments. I love you and hope that you grant me your satisfaction as always, because without it the blessing will be omitted from my life. Thanks to my father-in-law and my mother-in-law; Prof. Hossam Shalaby and Eng. Hoda El-Kaim for their continuous support, care and prayers. Special thanks to Prof. Hossam Shalaby for his unpaid help reviewing the thesis chapter by chapter to speed up my progress in writing the thesis.

Words are not enough and pages of this thesis will not be enough either to express my gratitude to my wife Mariam Shalaby. I will never forget her sacrifice with her career and leaving her family and friends in Egypt to accompany me in my life journey during this Ph.D. It is not just that she stood by me. She did bear the ups and downs with associated stress supporting me with love, care, and countless emotions. Indeed, her share of success in this thesis is more than mine. Thanks a lot for being my wife and thanks a lot for being my life-long companion. Also, I would like to thank my lovely girl; Janna Sowaillem. Your existence changed my life from sadness to happiness, from grief to hope. Your birth was the start of a new Mohammed Sowaillem. I am really blessed that you are my daughter. I am sorry that I was not able to spend more time with you. I hope that I will be a good father who you deserve. Your smile like your mother's gives me positive power to face the difficulties and cheer for accomplishments.

Many thanks to my brother; Moustafa Sowailem, and my sisters; Nesreen Sowailem, and Nancy Sowailem for their love support and prayers. Sorry for not being there, when you needed me during your life events. My apologies are to my niece, Farida Adel, and nephew, Omar Adel, whom I did not see until the moment of writing these words. I hope that time comes that I can express my love to you in actions not only words.

Many thanks to all my friends scattered all over the world. Thanks to my friends in Montreal for their support. Special thanks to Dr. Yasser Hifny for his help during the hardest period in my stay in Canada. My apologies continue to all who I was not able to mention by name, but I remember you in my heart and in my prayers.

In the end, this thesis, with all its worth, is dedicated to my wife, my parents and my two angels: Hamza and Janna.

Mohammed Sowailem

February 2017

Associated Publications

The outcome of the original contributions of the work associated with thesis is three published journal papers and three conference papers. In addition to two submitted papers in Photonics journal. The list of the publications is detailed below showing the contribution of the co-authors underneath each paper. In addition, I was a co-author in six journal and conference papers collaborating with other students. They are listed below as articles not related directly to this thesis.

Journal Articles Related to the Thesis

1. **M. Y.S. Sowailem**, T. M. Hoang, M. Morsy-Osman, M. Chagnon, D. Patel, S. Paquet, C. Paquet, I. Woods, O. Liboiron-Ladouceur, and D. V. Plant, “400-G Single Carrier 500-km Transmission With an InP Dual Polarization IQ Modulator,” *IEEE Photonics Technology Letters*, vol. 28, no. 11, pp. 1213-1216, June 2016.

Mohammed Sowailem wrote the paper. Mohammed Sowailem and Thang Hoang have equally contributed in building the test-bed for the InP device, optimizing the DSP code and other related experimental work. Mohammed Sowailem did the methodology for optimizing the digital pre-emphasis filter. Thang Hoang fitted the measured transfer function of the modulator to a mathematical function to apply on the signal samples. TeraXion (acquired by Ciena) provided the device, other measurements for device characterization and co-authors from TeraXion helped writing the device section. All co-authors helped either by fruitful discussions, transferring experimental knowledge and sharing some of the used code.

2. **M. Y.S. Sowailem**, T. M. Hoang, M. Morsy-Osman, M. Chagnon, M. Qiu, S. Paquet, C. Paquet, I. Woods, Q. Zhuge, O. Liboiron-Ladouceur, and D. V. Plant, “770 Gb/s PDM-32QAM Coherent Transmission Using InP Dual Polarization IQ Modulator,” *IEEE Photonics Technology Letters*, vol. 29, no. 5, pp. 442-445, March 2017.

Mohammed Sowailem wrote the paper. Mohammed Sowailem conceived the idea of using real valued MIMO with some adjustment to existing real valued MIMO algorithm. Mohammed Sowailem did the methodology for optimizing the digital pre-emphasis filter. Mohammed Sowailem built most of the setup with Thang Hoang help. Thang Hoang fitted the measured transfer function of the modulator to a mathematical function to apply on the signal samples. TeraXion (acquired by Ciena) provided the device. All co-authors helped

either by fruitful discussions, transferring experimental knowledge and sharing some of the used code.

3. **M. Y.S. Sowailem**, T. M. Hoang, M. Morsy-Osman, M. Chagnon, M. Qiu, S. Paquet, C. Paquet, I. Woods, O. Liboiron-Ladouceur, and D. V. Plant, “100G and 200G Single Carrier Transmission over 2880 and 320 km Using an InP IQ Modulator and Stokes Vector Receiver,” *Optics Express*, vol. 24, no. 26, pp. 30485-30493, Dec. 2016.

Mohammed Sowailem wrote the paper. Mohammed Sowailem and Thang Hoang have equally contributed in building the test-bed for the InP device, optimizing the DSP code and other related experimental work. Mohammed Sowailem wrote the core DSP code of the Stokes receiver and did the methodology for optimizing the digital pre-emphasis filter. Thang Hoang fitted the measured transfer function of the modulator to a mathematical function to apply on the signal samples and did measurements related to optimizing the carrier to signal power ratio which is a key parameter in such system. Ciena provided the device, other measurements for device characterization. All co-authors helped either by fruitful discussions, transferring experimental knowledge and sharing some of the used code.

4. **M. Y.S. Sowailem**, T. M. Hoang, M. Morsy-Osman, M. Chagnon, M. Qiu, S. Paquet, C. Paquet, I. Woods, O. Liboiron-Ladouceur, and D. V. Plant, “Impact of Chromatic Dispersion Compensation in Single Carrier Two Dimensional Stokes Vector Direct Detection System,” submitted to *IEEE Photonics Journal*.

Mohammed Sowailem conceived the idea, wrote the paper and wrote the DSP code. Mohammed Sowailem, Thang Hoang, and Mathieu Chagnon have built the experimental setup. Mohammed Sowailem did most of the experimental work with the help of Thang Hoang. Mohammed Sowailem did the mathematical analysis verified by Mohamed Osman. Ciena provided the device, other measurements for device characterization. All co-authors helped either by fruitful discussions, transferring experimental knowledge and sharing some of the used code.

5. **M. Y.S. Sowailem**, E. El-Fiky, M. Morsy-Osman, Q. Zhuge, T. M. Hoang, S. Paquet, C. Paquet, I. Woods, O. Liboiron-Ladouceur, and D. V. Plant, “Self-homodyne system for next generation datacenter interconnects in intra-datacenters,” submitted to *Optics Express*.

Mohammed Sowailem conceived the idea, wrote the paper and wrote the DSP code. Mohammed Sowailem has built the experimental setup. Mohammed Sowailem did most of the experimental work. Eslam EL-Fiky helped in some stages of the experimental work. Ciena provided the device, other measurements for device characterization. All co-authors helped either by fruitful discussions, transferring experimental knowledge and sharing some of the used code

Conference Articles Related to the Thesis

1. **M. Y.S. Sowailem**, M. Morsy-Osman, O. Liboiron-Ladouceur and D. V. Plant, “A self-coherent system for short reach applications,” in *Proc. Photonics North (PN) 2016*, Quebec City, QC, 2016, pp. 1-1.

Mohammed Sowailem conceived the idea, did the simulation and wrote the paper. All co-authors helped either by fruitful discussions, transferring simulation knowledge and sharing some of the used code

2. **M. Y.S. Sowailem**, D. V. Plant, and O. Liboiron-Ladouceur, “Contention resolution strategy in optical burst switched datacenters,” in *Proc. Optical Fiber Communication Conference and Exposition and the National Fiber Optic Engineers Conference (OFC/NFOEC) 2013*, pp.1-3, Anaheim, CA, USA, 17-21 March 2013.

Mohammed Sowailem conceived the idea, did the simulation and wrote the paper. All co-authors helped with fruitful discussions.

3. **M. Y.S. Sowailem**, D. V. Plant, O. Liboiron-Ladouceur, “Implementation of optical burst switching in data centers,” *IEEE Photonics Conference (PHO) 2011*, Arlington, VA, 2011, pp. 445-446, 9-13 Oct. 2011.

Mohammed Sowailem conceived the idea, did the simulation and wrote the paper. All co-authors helped with fruitful discussions.

Journal Articles Not Directly Related to the Thesis

1. E. El-Fiky, M. Chagnon, **M. Sowailem**, A. Samani, M. Morsy-Osman and D. V. Plant, “168-Gb/s Single Carrier PAM4 Transmission for Intra-Data Center Optical Interconnects,” in *IEEE Photonics Technology Letters*, vol. 29, no. 3, pp. 314-317, Feb. 2017.

2. M. Qiu, Q. Zhuge, **M. Y.S. Sowailem**, T. Hoang, M. Chagnon, M. Xiang, X. Zhou, F. Zhang, and D. V. Plant, "Equalization-Enhanced Phase Noise in Stokes-Vector Direct Detection Systems," *IEEE Photonics Journal*, vol. 8, no. 6, pp. 1-7, Dec. 2016.
3. T. Hoang, **M. Sowailem**, M. M. Osman, M. Chagnon, D. Patel, C. Paquet, S. Paquet, I. Woods, O. Liboiron-Ladouceur, and D. Plant, "Transmission of 344 Gb/s 16-QAM using a simplified coherent receiver based on single-ended detection", *IEEE Photonics Journal*, vol.8, no. 3, pp. 1-8, June 2016.
4. T. Hoang, M. M. Osman, M. Chagnon, M. Qiu, D. Patel, **M. Sowailem**, X. Xu and D. Plant , "Phase-diversity method using phase-shifting interference algorithms for digital coherent receivers", *Optics Communications*, vol. 356, pp. 269-277, Dec. 2015.
5. W. Wang, Q. Zhuge, Y. Gao, M. Qiu, M. Chagnon, **M. Y. Sowailem**, F. Zhang, D. V. Plant, "Enhanced Channel Equalizers for Adaptive Zero-Guard-Interval CO-OFDM Systems," *IEEE Photonics Technology Letters*, vol. 27, no. 16, pp. 1721-1724, Aug. 2015.

Conference Articles Not Directly Related to the Thesis

6. O. Liboiron-Ladouceur, M. N. Sakib, **M. Sowailem**, M. S. Hai, R. Varano, and D. Rolston, "Optical Interconnected High-performance Servers", *SPIE Photonics North 2012*, Montreal, QC, June 2012.

Contents

Abstract	i
Résumé	iii
Acknowledgments	vi
Associated Publications	ix
Contents	xiii
List of Figures	xvi
List of Tables	xviii
List of Acronyms	xix
Chapter 1	1
Introduction	1
1.1 Motivation	1
1.2 Background summary and literature review	3
1.2.1 Optical switching for high speed optical systems in intra-Datacenter networking	3
1.2.2 High speed integrated optical solutions for datacenter interconnects in intra-datacenter networking	7
1.2.3 High speed integrated optical systems for datacenter interconnects in inter-datacenter networking	8
1.3 Problem statement and thesis objectives	13
1.4 Original contribution	14
1.5 Thesis organization	16
Chapter 2	18
Optical Burst Switching in Intra-Datacenter Networking	18
2.1 Overview	18
2.2 Network architecture approaches for OBS in intra-datacenter networking	22
2.2.1 Active approach for OBS network architecture in intra-datacenter networking	22
2.2.2 Passive approach for OBS network architecture in intra-datacenter networking using AWGR-based switching mechanism	23

2.3	Hybrid burst assembly algorithm.....	25
2.4	Self-similarity and burst assembly	27
2.5	Control plane layer for OBS in intra-datacenter networking	29
2.5.1	Resource allocation for new connection requests	30
2.5.2	Contention resolution and OBF algorithm.....	32
2.6	Simulation results and discussion.....	33
2.6.1	Active approach simulation assumptions, results and discussion.....	33
2.6.2	Passive approach simulation assumptions, results and discussion	35
2.7	Conclusion	40
Chapter 3		42
Self-Homodyne System for Datacenter Interconnects in Intra-Datacenter Networking.....		42
3.1	Overview	42
3.2	Principle of the proposed self-homodyne system.....	44
3.3	Experimental setup	46
3.4	DSP stack.....	47
3.5	Results and discussion	49
3.6	Conclusion	53
Chapter 4		55
High Speed Coherent Optical System Using InP Integrated Modulator		55
4.1	Overview	55
4.2	Description of the deployed InP-based DP-IQM	57
4.3	Experimental setup	58
4.3.1	Experimental setup for 400G coherent transmission demonstration	58
4.3.2	Experimental setup for 600G coherent transmission demonstration	60
4.4	DSP stack	61
4.4.1	DSP stack used for 400G coherent transmission	61
4.4.2	DSP stack used for 600G coherent transmission	62
4.5	Results and discussion	64
4.5.1	400G coherent transmission demonstration results	64
4.5.2	600G coherent transmission demonstration results	68
4.6	Conclusion	71
Chapter 5		72
Stokes Vector Direct Detection System for Inter-Datacenter Interconnect Applications ...		72
5.1	Overview	72
5.2	Principle of 2D-SV-DD system and CD limitations	75
5.2.1	Principle of 2D-SV-DD system transceiver	75
5.2.2	CD limitations in 2D-SV-DD system	76

5.3	Experimental setup	78
5.3.1	First phase of 2D-SV-DD system demonstration	78
5.3.2	Second phase for CD impact on 2D-SV-DD demonstration	80
5.4	DSP stack	82
5.4.1	First phase of 2D-SV-DD system DSP stack.....	82
5.4.2	Second phase DSP stack for CD impact on 2D-SV-DD.....	84
5.5	Results and discussion	86
5.5.1	First phase of 2D-SV-DD system demonstration	86
5.5.2	Second phase for CD impact on 2D-SV-DD demonstration	90
5.6	Conclusion	92
Chapter 6	94
Conclusion and Future Work	94
6.1	Summary of presented work.....	94
6.2	Future research avenues.....	96
6.2.1	Potential research extension for presented work	96
6.2.2	Roadmap for future implementation	98
Bibliography	100

List of Figures

Fig. 1.1 AWGR wavelength routing feature where $\lambda w(i,o)$ means wavelength number w connects between input number i and output number o .	6
Fig. 1.2 AWGR cyclic property periodicity.	6
Fig. 1.3 Coherent transmission system architecture.	9
Fig. 1.4 Basic DSP stack: (a) Tx-DSP, and (b) Rx-DSP.	11
Fig. 1.5 Schematic for 2D configuration for SV-DD system.	13
Fig. 2.1 Active approach for OBS network architecture: (a) Main intra-datacenter network topology with centralized scheduling; (b) pod interface to core layer switching; (c) core switch bank based on active fast fabric switches.	23
Fig. 2.2 Passive approach for OBS network architecture: (a) Main intra-datacenter network topology with centralized scheduling (not connected to core layer); (b) pod interface to core layer switching; (c) AWGR-based switching mechanism with E/O and O/E block description.	25
Fig. 2.3 Hybrid Burst Assembly Algorithm.	26
Fig. 2.4 Probability Distribution function for lognormal distribution at different shape parameters.	28
Fig. 2.5 Success probability versus mean packet inter-arrival times for different buffer limits (no contention resolution, infinite buffer limit and finite buffer limit).	34
Fig. 2.6 Latency versus mean packet inter-arrival times (latency due to single burst construction is common for all buffer limits, latency due to contention resolution for different buffer limits).	34
Fig. 2.7 Buffer utilization versus mean packet inter-arrival times for different buffer limits.	35
Fig. 2.8 Normalized throughput versus packet offered load for both burst and packet switching in both cases; exponential case and lognormal case (at different degrees of self-similarity).	36
Fig. 2.9 Latency versus packet offered load for both burst and packet switching in both cases; exponential case and lognormal case (at different degrees of self-similarity).	37
Fig. 2.10 Normalized throughput versus packet offered load at different buffer size limits.	38
Fig. 2.11 Latency versus packet offered load at different buffer size limits.	39
Fig. 2.12 Normalized average throughput (average over all switches going to a single pod and normalized w.r.t. data rate) versus normalized average traffic load offered from a single pod.	40
Fig. 3.1 Schematic of the proposed self-homodyne system.	45
Fig. 3.2 Experimental setup for the proposed system emulation.	46
Fig. 3.3 DSP stack for a self-homodyne system; (a) Tx-DSP, and (b) Rx-DSP.	49
Fig. 3.4 BER versus distance at different symbol rates using 16QAM modulation format.	49
Fig. 3.5 56 Gbaud 16QAM spectrum from the captured data at the receiver.	50
Fig. 3.6 BER versus distance at different symbol rates using 32QAM modulation format.	51
Fig. 3.7 BER versus path mismatch length between the tone and the modulated signal using phase noise mitigation and without using phase noise mitigation.	52
Fig. 3.8 BER versus receiver number of taps used in the real valued 4×4 MIMO at different number of taps used in the digital pre-emphasis at the transmitter.	53
Fig. 4.1 InP-based DP-IQM schematic (CB-CPW: Conduction-backed coplanar waveguide).	58
Fig. 4.2 Experimental setup used for 400G coherent transmission demonstration.	59
Fig. 4.3 Experimental setup used for 600G coherent transmission demonstration. Inset: B2B optical spectrum with 0.05 nm resolution bandwidth for 84 Gbaud signal with digital pre-emphasis (red curve) and without digital pre-emphasis (blue curve).	60
Fig. 4.4 DSP stack used for 400G coherent transmission demonstration: (a) Tx-DSP; (b) Rx-DSP.	62
Fig. 4.5 DSP stack used for 600G coherent transmission demonstration: (a) Tx-DSP for 84 Gbaud at one sample per symbol (sps); (b) Tx-DSP for 77 Gbaud; (c) Rx-DSP.	63
Fig. 4.6 BER versus distance at 56 Gbaud 16QAM per launched power in 400G coherent transmission demonstration with B2B constellation as an inset.	65
Fig. 4.7 BER versus OSNR for 56 Gbaud 16QAM in 400G coherent transmission demonstration.	66
Fig. 4.8 BER versus distance for 38 Gbaud and 43 Gbaud 64QAM in 400G coherent transmission demonstration with B2B constellations as an inset.	67
Fig. 4.9 BER versus OSNR for 38 Gbaud 64QAM system performance as a function of modulator platform.	68

Fig. 4.10 BER versus distance for 77 Gbaud PDM-32QAM signal, 84 Gbaud PDM-16QAM, PDM-8QAM and PDM-QPSK signals for two cases; using 2×2 complex MIMO in Rx-DSP and using 4×4 real valued MIMO in Rx-DSP	69
Fig. 4.11 Payload rate after overall overhead removal versus achieved reach (Note: overall overhead 28% in case of SD-FEC). Inset: B2B constellations.	69
Fig. 4.12 BER versus OSNR for 77 Gbaud 32QAM signal, 84 Gbaud 16QAM and QPSK signals. Dotted lines represent B2B BER floors for each scenario which are taken from the BER curves at 0 km.	70
Fig. 5.1 Schematic diagram of the 2D-SV-DD system.	75
Fig. 5.2 First phase of 2D-SV-DD demonstration (a) Experimental setup (b) SVR details (c) B2B optical spectrum with 0.05 nm resolution bandwidth.	79
Fig. 5.3 Second phase of 2D-SV-DD demonstration experimental setup showing the following insets: (i) SVR details, (ii) recirculating loop details, (iii) B2B optical spectrum with 0.05 nm OSA resolution bandwidth in case of the 84 Gbaud signal, and (iv) RTOs configuration according to the operating symbol rates.	81
Fig. 5.4 DSP stack used for the first phase of 2D-SV-DD demonstration: (a) Tx-DSP; (b) Rx-DSP.	83
Fig. 5.5 Transmitter and receiver DSP stacks for two CD compensation methods: (a) CD pre-compensation, and (b) CD post-compensation. The inset on the right shows one long Stokes training pulse in B2B and after 320 and 640 km in case of a 84 Gbaud signal.	84
Fig. 5.6 BER versus transmission distance at the optimum CSPR and the optimum launch power for QPSK, 16QAM, and 8QAM modulation formats at 56 Gbaud using 2D-SV-DD.	87
Fig. 5.7 BER versus CSPR of various format at fixed transmitted distance with constellations.	88
Fig. 5.8 BER versus distance (CD pre-compensation and post-compensation comparison at 84 Gbaud and 64 Gbaud 16QAM cases).	91
Fig. 5.9 BER versus CSPR for 84 Gbaud 16QAM and 64 Gbaud 16QAM using CD pre-compensation.	91
Fig. 5.10 BER versus distance in case of CD pre-compensation for signals with symbol rates of 64 Gbaud and 84 Gbaud at different modulation formats	92

List of Tables

Table 2-1 Hurst parameter of traffic at different scenarios.....	37
Table 5-1 Schemes for SV-DD compared to 400G coherent system demonstration (SP-IQM: single polarization IQ modulator, IFFT: inverse fast Fourier transform).....	89

List of Acronyms

2D-SV-DD	Two-Dimensional Stokes Vector Direct Detection
ADC	Analog-to-Digital Converter
ASE	Amplified Spontaneous Emission
AWGN	Additive White Gaussian Noise
AWGN	Additive White Gaussian Noise
AWGR	Arrayed Waveguide Grating Router
B2B	Back-to-Back
BER	Bit Error Rate
BPD	Balanced Photodetector
CAGR	Compound Annual Growth Rate
CAP	Carrierless Amplitude Phase
CD	Chromatic Dispersion
CFP	Common Form-factor Pluggable
CMOS	Complementary Metal Oxide Semiconductor
CB-CPW	Conduction-Backed Coplanar Waveguide
CRx	Coherent Receiver
CSPR	Carrier-to-Signal Power Ratio
CW	Continuous Wave
CWDM	Coarse Wavelength-Division Multiplexing
DAC	Digital-to-Analog Converter
DCI	Datacenter Interconnect

DD	Direct Detection
DD-LMS	Decision Directed Least Mean Squares
DML	Directly Modulated Laser
DMT	Discrete Multi-Tone
DP	Dual Polarization
DP-IM	Dual Polarization Intensity Modulation
DP-IQM	Dual Polarization IQ Modulator
DSP	Digital Signal Processing
DWDM	Dense Wavelength-Division Multiplexing
E/O	Electrical-to-Optical
ECL	External Cavity Laser
EDFA	Erbium Doped Fiber Amplifier
ENOB	Effective Number of Bits
ETDM	Electrical Time-Division Multiplexing
FCFS	First Come First Served
FDL	Fiber Delay Line
FEC	Forward Error Correcting
FIFO	First-In-First-Out
FIR	Finite-Impulse-Response
FSR	Free Spectral Range
FWM	Four Wave Mixing
HD-FEC	Hard Decision Forward Error Correcting
HOL	Head of Line
IFFT	Inverse Fast Fourier Transform

IM	Intensity Modulation
IM/DD	Intensity Modulation/Direct-Detection
InP	Indium Phosphide
IoT	Internet of Things
IQ	In-phase/quadrature-phase
ISI	Inter-Symbol Interference
JET	Just-Enough-Time
JIT	Just-In-Time
LO	Local Oscillator
LOBS-HC	Labeled Optical Burst Switching with Home Circuits
LQF	Longest Queue First
LSPR	LO-to-Signal-Power-Ratio
MEMS	Micro-Electro-Mechanical System
MIMO	Multiple-Input-Multiple-Output
MZM	Mach-Zehnder Modulator
O/E	Optical-to-Electrical
OBF	Oldest Burst First
OBS	Optical Burst Switching
OCF	Oldest Cell First
OCS	Optical Circuit Switching
OFDM	Orthogonal Frequency Division Multiplexing
OOK	On-Off Keying
OPS	Optical Packet Switching
OSA	Optical Spectrum Analyzer

OSNR	Optical Signal-to-Noise Ratio
OTDM	Optical Time-Division Multiplexing
PAM-4	4-level Pulse Amplitude Modulation
PAPR	Peak-to-Average-Power Ratio
PBC	Polarization Beam Combiner
PBS	Polarization Beam Splitter
PC	Polarization Controller
PD	Photodetector
PDF	Probability Density Function
PDM	Polarization-division Multiplexing
PICs	Photonics integrated circuits
PLL	Phase Locked Loop
PM	Polarization Maintained
PMD	Polarization Mode Dispersion
POD	Performance-Optimized Datacenters
QAM	Quadrature Amplitude Modulation
QoS	Quality of Service
QPSK	Quadrature Phase Shift Keying
RC	Raised Cosine
RRC	Root Raised Cosine
RTO	Real Time Oscilloscope
Rx-DSP	Receiver Digital Signal Processing
SBS	Stimulated Brillouin Scattering
SD-FEC	Soft Decision Forward Error Correcting

SDN	Software-Defined-Network
SDRAM	Synchronous Dynamic Random Access Memory
SE-PD	Single-Ended PD
SiP	Silicon Photonics
SMF	Single Mode Fiber
SNR	Signal-to-Noise Ratio
SOA	Semiconductor Optical Amplifier
SOP	State of Polarization
SP	Single Polarization
SP-IQM	Single Polarization IQ Modulator
SP-IQM	Single Polarization IQ Modulator
SPM	Self Phase Modulation
SRS	Stimulated Raman Scattering
SSMF	Standard Single Mode Fiber
SVR	Stokes Vector Receiver
TAG	Tell-And-Go
TDM	Time-Division Multiplexing
TEC	Thermoelectric Cooler
TIA	Transimpedance Amplifiers
TL	Tunable Laser
TS	Training Symbols
TS-LMS	Training Symbol Least Mean Squares
TWC	Tunable Wavelength Converter
Tx-DSP	Transmitter Digital Signal Processing

ULAF	Ultra Large Area Fiber
VCSELs	Vertical Cavity Surface Emitting Lasers
VOA	Variable Optical Attenuator
VODL	Variable Optical Delay Line
VOQ	Virtual Output Queue
WAN	Wide Area Network
WDM	Wavelength-Division Multiplexing
XPM	Cross Phase Modulation
ZB	Zettabytes

Chapter 1

Introduction

1.1 Motivation

The tremendous increase in global IP traffic, driven by cloud computing, Internet of Things (IoT), and data storage, reflects on capacity demand in the data-communication industry. According to the Cisco global cloud index forecast [1], the global datacenter IP traffic will experience more than three-fold increase in the period from 2015 to 2020 to reach almost 15.3 Zettabytes (ZB) ($1 \text{ ZB} = 10^{21}$ bytes) by 2020. Out of the global datacenter traffic, about 77% is concentrated in the intra-datacenter networking, i.e., it stays within the same datacenter. This large percentage of intra-datacenter traffic can be explained in light of the huge amount of background data processing and associated applications, which generates much more traffic than what is delivered to the end users. Since intra-datacenter interconnects have traditionally relied on either electrical interconnects over copper links or optical interconnects over multimode fibers and utilizing vertical cavity surface emitting lasers (VCSELs), these current solutions can no longer provide the required capacity or reach (due to increasing datacenter sizes) [2]. Hence, in order to cope with the above-mentioned massive intra-datacenter traffic growth, there is an immense need for next generation low-cost and power efficient optical interconnects operating over single mode fiber to deliver higher capacities ($> 100 \text{ Gb/s}$) over typical intra-datacenter reaches of up to 10 km [2].

Along with the need for high speed optical interconnects operating over single mode fibers for intra-datacenter interconnects, the capacity increase of these transceivers, that are traditionally plugged into the faceplates of core/aggregate switches, will push the overall switching capacity to levels at which electronic switching is no longer a viable solution. Alternatively, all-optical or hybrid electronic/optical switching must replace conventional electronic switching in order to provide the desired future switching capacity while maintaining a reasonable power consumption [2]. In all-optical switching, the reduction in power consumption is attributed to the absence of optical-to-electrical (O/E) and electrical-to-optical (E/O) conversion at the core layer electronic switches. However, the optical switching technology still lacks maturity especially

with the absence of a practical methodology to resolve potential contentions by buffering packets all-optically. Along with the above challenges, another important aspect in intra-datacenter networking is the massive number of connections required by DCIs, which pushes the demand towards having small form factor solutions to enable compact pluggable interconnects and provide the high port density required. Photonics integrated circuits (PICs) provide promising solutions for integrated pluggable solutions where several optical components can be integrated on one host material beside the potential co-integration with CMOS electronic devices. Silicon photonics (SiP) and Indium Phosphide (InP) platforms are two candidates capable of hosting integrated solutions for DCIs [3-6].

In parallel with the escalating traffic volume in intra-datacenter applications, there is also a growing capacity demand for inter-datacenter links, which interconnect geographically distant datacenters. According to the Cisco global cloud index forecast [1], the compound annual growth rate (CAGR) of inter-datacenter traffic over the five-year period from 2015 to 2020 is nearly 32%, which is the highest compared to CAGR of intra-datacenter and datacenter-to-user traffic. This highlights the importance of addressing the increasing capacity demand over inter-datacenter reaches, which are typically within few hundreds of kilometers. Over these reaches, we face additional challenges besides those encountered over intra-datacenter reaches. For example, optical propagation loss is hardly an issue for short-reach intra-datacenter interconnects because of the short propagation distance whereas for inter-datacenter reaches, propagation loss is a major impairment that affects the signal quality. Thus, optical amplifiers must be employed and C-band operation around 1550 nm is typically adopted due to the lower propagation loss coefficient of the optical fiber at this operating wavelength in addition to the availability of mature optical amplification technology by means of Erbium doped fiber amplifiers (EDFAs). Since inter-datacenter DCIs should operate in the C-band, they have to be capable of compensating fiber chromatic dispersion (CD). In addition, they must provide high spectral-efficiency to achieve the desired aggregate capacity per link. Optical coherent system is a potential solution that achieves both requirements. The coherent detection at the receiver allows the full compensation of CD with digital signal processing, i.e., without the use of optical dispersion compensation which would add insertion loss to the total link loss if employed [7]. In addition, coherent transmission provides very high spectral efficiency with the use of advanced modulation formats, e.g., quadrature amplitude modulation (QAM) compared to On-Off keying

(OOK) adopted in legacy intensity modulation/direct-detection (IM/DD) transceivers. However, the cost of implementing a coherent transceiver is relatively higher than direct-detection schemes, which is typically not a concern for long-haul/metro applications but may prove prohibitive for implementation over some of inter-datacenter links. Hence, there is a need for alternative solutions that are: 1) capable of providing spectral efficiency higher than IM/DD, 2) able to combat chromatic dispersion, and 3) more cost efficient relative to coherent.

1.2 Background summary and literature review

In this section, we present an essential summary for the technical background of the topics covered in this thesis and the related literature review. In subsection 1.2.1, we introduce the concept of the optical burst switching and different physical realization methods for optical switching. In addition, we present approaches available in literature to implement optical switching in the intra-datacenter networking. Subsection 1.2.2 is dedicated to highlight the challenges associated with the DCIs in intra-datacenter networking and different approaches to meet those challenges. Finally, we continue addressing the DCIs, but in the inter-datacenter networking applications. In subsection 1.2.3.1, we provide a background to the optical coherent system used in inter-datacenter communication, and then we describe another potential solution in literature using Stokes vector direct detection system in subsection 1.2.3.2.

1.2.1 Optical switching for high speed optical systems in intra-Datacenter networking

1.2.1.1 Optical burst switching

Optical burst switching (OBS) network architecture consists of three main network components: ingress node (source), egress node (destination), and core node. In the ingress node, the aggregation of the data packets into data bursts (the transmission unit in OBS) in assembly buffers takes place according to one of the assembly algorithms [8]. The three main assembly algorithms in the OBS are burst-length-based, time-based, and hybrid burst assembly algorithms [8-10]. Thus, the data burst has variable length that depends on the packet offered load and the threshold or thresholds of the used assembly algorithm.

One of the main features of the OBS is the separation between the control plane and the data plane. The control packet is sent prior to the data burst on dedicated channels for the control

packets to configure the core nodes in the path [8, 11, 12]. There are three main types for signaling protocol:

- Centralized signaling protocol which requires the existence of the ingress nodes, core nodes, and the egress nodes in a limited geographical area [8, 11].
- Two-way reservation protocol where the control packet is sent configuring the core nodes in the path between the ingress node and the egress node, meanwhile the ingress node waits for a response to send the data burst [8].
- One-way reservation protocol where the control packet is sent followed by the data burst either immediately or with an offset. If the data burst is sent immediately after the control packet the data packet needs to be delayed using a fiber delay line (FDL) until the control packet is processed. This signaling protocol is called tell-and-go (TAG) signaling protocol [8]. If there is an offset between the control packet and the data burst, two other signaling protocols can be used namely; just-in-time (JIT), and just-enough-time signaling protocols [8]. In JIT, the core node reserve the required resources since the control packet arrival until it receives a release packet from the ingress node [13]. In JET, the control packet contains information about the burst length and the time offset between the control packet and the data burst [14]. The core node reserve the resources for duration equivalent to the data burst length, when the data burst reaches the core node.

Finally, the core node uses one of the scheduling algorithms to assign a wavelength to the data burst. The used scheduling algorithm can depend on the voids (idle periods between the scheduled data bursts) or can depend only on the availability in the horizon. More details about the proposed scheduling algorithm for OBS can be found in [8, 10, 12].

1.2.1.2 Optical switching physical realization

For all-optical switching physical realization, there are two methods; either the broadcast-selective method or the space switching method. In the broadcast-selective method, the optical signal at each input port is split into multiple branches; each branch is dedicated to one of the output ports. After selection by ON-OFF elements like semiconductor optical amplifiers (SOAs), all branches that are dedicated to a certain output port are combined at this output port [15]. The

problem of this method is the degradation of the optical signal due to excessive power loss associated with the broadcast process, which limits the number of ports. In [16], the authors used EDFA to compensate for the power loss. On the other hand, the space switching method connects between an input/output pair without power splitting. There are different approaches for this method implementation like:

- The use of MEMS. However, the configuration time of MEMS is relatively high, i.e. approximately 3-10 msec [17]. In [18], MEMS are used when appropriate along with electronic switching.
- The use of MZI in an optical-cross-connect switch [19]. This implementation causes signal degradation and the realized switches are with small number of ports [20, 21].
- The use of AWGR-based switching mechanisms [22-28]. This approach is scalable and signal quality at the output ports is sufficient for reception excluding cases where the signal passes through the AWGR several times as in [23], where authors need to use EDFA in some cases when they resolve the contention using fiber delay lines.

The most important properties of the AWGR switching element can be summarized as follows:

- The wavelength routing property: This property implies that the destined output port is dependent on the wavelength used at a given input port. The AWGR allows any input port to be connected to any output port using the appropriate wavelength as shown in Fig. 1.1.
- The cyclic property: This property depends on a spectral range known as the free spectral range (FSR) where a wavelength, that is used to connect a given input-output port pair, can connect between other pairs in a cyclic fashion. Fig. 1.1 illustrates this cyclic property where, for example, inputs $i = 1, 2,$ and 3 can communicate simultaneously with outputs $o = N, 1,$ and $2,$ respectively, at same wavelength λ_N .
- The periodicity of the cyclic property every FSR where within each FSR includes a number of wavelengths equal to the number of ports and the cyclic property is effective among these wavelengths. For example, if the number of ports is N , the first FSR contains N wavelengths. Each input port can use these wavelengths to communicate with the N output ports. The second FSR contains N wavelengths that

can be used in the same fashion as wavelengths in the previous FSR as illustrated in Fig. 1.2, which is taken from [27].

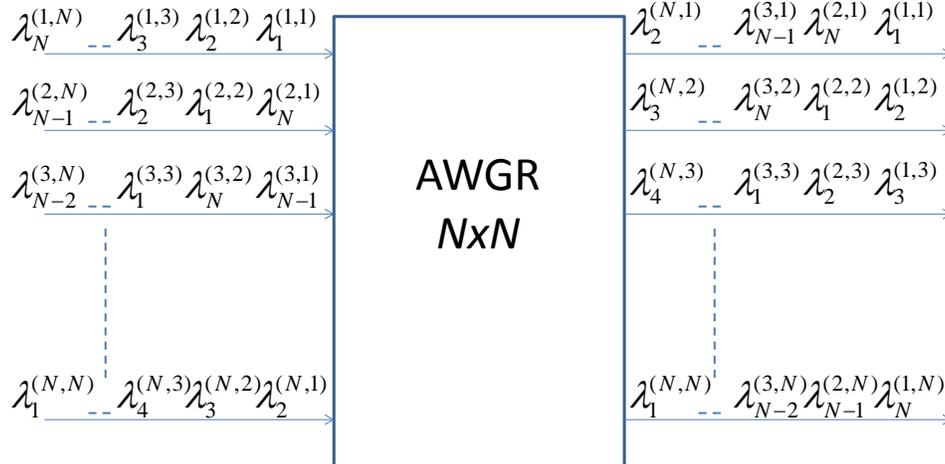


Fig. 1.1 AWGR wavelength routing feature where $\lambda_w^{(i,o)}$ means wavelength number w connects between input number i and output number o .

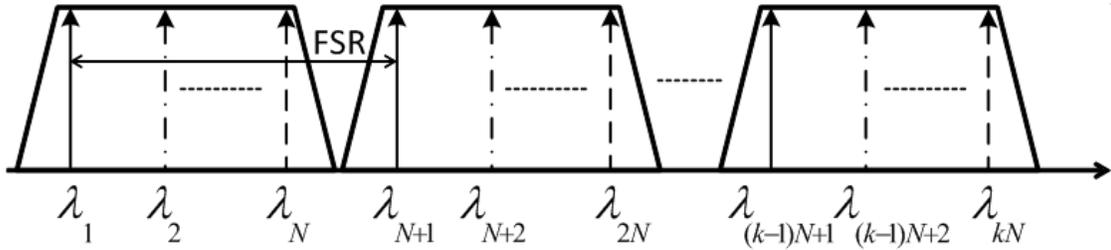


Fig. 1.2 AWGR cyclic property periodicity.

1.2.1.3 Optical switching implementation in intra-datacenter networking and contention resolution strategies

As mentioned in section 1.1, the use of optical switching is necessary not only to meet current challenges, but also to provide a reliable solution for next generation networks in datacenters. Thus, all-optical switching implementation in intra-datacenter networking and high performance computing networks has become one of the cutting-edge technologies that concerns researchers in datacom [22, 23, 26, 29-33].

One of the proposed solutions for optical switching is to use a single AWGR equipped with tunable wavelength converters (TWCs) at its input ports for switching [22]. Optical label switching is used, where the optical label is extracted and forwarded to a central control unit. The

data packet is optically delayed in a FDL until the label is processed; scheduling is done and the TWC is configured by the control unit. If contention occurs, packets are buffered in a synchronous dynamic random access memory (SDRAM) after O/E conversion. When the contention is over, packets are retransmitted exhibiting another E/O conversion.

Similarly in [23], an AWGR, equipped with TWCs at its input and output ports, is used as a single switch in a multistage switching system. FDLs are used instead of SDRAM for contention resolution. EDFA are used to make the signal level suitable as an input to the TWC. The TWC is adjusted such that the optical signal is kept circulating in the FDLs until the contention is resolved.

In [33], authors used centralized control along with the burst assembly buffers to resolve the contention. Horizon scheduling algorithm is used based on a synchronous system to schedule the burst transmission based on prospective connection releases [8]. The centralized scheduler assigns the wavelength and the start time for burst transmission for the source pod. Although, the authors provided options for switches that can be used in the network architecture, they did not address the physical implementation of the optical switching and how it interacts with the control plane.

1.2.2 High speed integrated optical solutions for datacenter interconnects in intra-datacenter networking

As mentioned in section 1.1, the intra-datacenter networking where cloud services take place has the largest share of IP traffic. Thus, the implementation of high speed and cost effective DCIs is necessary not only for current data rate requirements, but also for next generation intra-datacenter networking that requires 400G and beyond transmission. 100 Gb/s transmissions for short reach application has already been standardized [34]. The main idea is to have four fiber lanes or four wavelengths; each carrying a 25 Gb/s OOK signals. In the 400 GbE task force [35], a more advanced modulation format; namely 4-level pulse amplitude modulation (PAM-4), is adopted in two different configurations. First, four fiber lanes operating at a symbol rate of 50 Gbaud is the configuration chosen for Ethernet frames transmission over 500 m of SMF links. The second configuration uses eight wavelengths on 800 GHz grid; each carrying 25 Gbaud PAM-4 signal, for transmission over 2 km and 10 km. However, capacity per wavelength shortly requires a significant increase to achieve 800 Gb/s and 1.6 Tb/s transmission [36].

Recently, high-speed experimental demonstrations for IM/DD transmission have been reported using various advanced modulation formats. The most common approach is increasing the symbol rate using PAM-4 modulation format [37-44]. Another proposed technique is the half-cycle Nyquist subcarrier modulated 16-ary quadrature amplitude modulation (16QAM) with polarization-division multiplexing (PDM) [45, 46]. However, the polarization demultiplexing was done using manual adjustment of the polarization controller. Similar approach called multi-band carrierless amplitude phase modulation (multi-CAP) allows for QAM modulation in a direct detection system [40, 47-49]. Finally, another well-known candidate for high-speed IM/DD transmission is the discrete multi-tone (DMT) [40, 50, 51]. For example, 400 Gb/s transmission has been experimentally demonstrated as 4×100 Gb/s using DMT and 4 directly modulated lasers (DMLs) over 30 km in [50] in the O-band.

Along with demonstrations for the future 400 Gbps and 1 Tbps short reach links, the technology of the used platform is investigated for integrated and cheap solutions for DCIs. SiP and InP devices are two candidates for integrated photonics platforms [3, 4]. Both are useful host material for fabricating photonic integrated devices (e.g., the modulator, photodiodes, etc.) because of large scale integration, and CMOS compatibility, enabling co-integration with SOA, lasers, and RF drive amplifiers [6, 52, 53]. Several demonstrations have been reported using SiP [37, 54-57] and InP devices [52, 53, 58-63] as potential integrated solutions in optical systems. InP modulators have the advantage of achieving high bandwidth at much lower V_π which may be as low as 1.5 V [63, 64]. This reduces the driving RF requirements which reduce the overall power consumption in large-scale implementation like in intra-datacenter networking.

1.2.3 High speed integrated optical systems for datacenter interconnects in inter-datacenter networking

1.2.3.1 Optical coherent systems

For intra-datacenter reaches (less than 40 km), intensity modulation at the transmitter and direct detection at the receiver (IM/DD) is adopted for cost efficient system and to enable small form factor devices. Although inter-datacenter networking has the same requirements in terms of deploying cost effective small form factor pluggable solutions, the inter-datacenter reaches (<1000 km) make the coherent detection be the current choice for inter-datacenter networking. In

this subsection, we present basics of optical coherent transmission and associated digital signal processing at the transmitter (Tx-DSP) and receiver (Rx-DSP).

In classical single carrier dual polarization optical coherent system, the electric fields of the X-polarization and Y-polarization of a continuous wave laser source are modulated at the transmitter such that each polarization carries independent complex symbols using optical in-phase/quadrature modulator (IQM) as shown in Fig. 1.3. The RF signals used by the IQM to modulate the optical signal are generated by four DACs followed by RF linear drivers. Next, the modulated optical signal is launched over the optical channel, which includes several spans of single mode fibers (SMFs). An inline EDFA follows each fiber span to compensate for the optical attenuation due to signal propagation in the SMF link. At the Rx, a coherent front-end shown in Fig. 1.3 uses a local oscillator (LO) with the same wavelength used by the laser source at the transmitter for coherent detection. Two polarization beam splitters (PBSs) are used to split X-polarization and the Y-polarization of the signal and the LO. Each polarization is introduced to a 90° optical hybrid followed by four balanced photodetectors (BPDs) to provide four signals corresponding to the I and Q components on both polarizations of the received optical field. The analogy of beating between the LO and the signal on the photodetectors after the 90° hybrid is the same sense as RF mixing between the local oscillator (sin and cos) and the wireless signal with addition to the low pass filter effect. However, in optical detection, we have direct detection terms which are the signal instantaneous power and the LO power. That is why we use BPDs to remove the direct detection terms. The output baseband signals from the balanced detection are then sampled by ADCs and processed by the Rx-DSP.

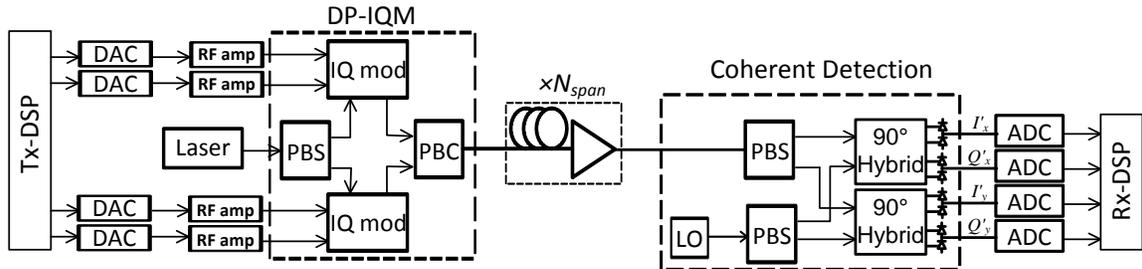


Fig. 1.3 Coherent transmission system architecture.

Before handling the DSP stack in the optical coherent system, we introduce the main system impairments affecting the used DSP algorithms. First, we review briefly impairments accompanied by the optical signal propagation in the optical fiber. Beside attenuation compensated by using the inline EDFAs, the optical signal undergoes:

- Chromatic dispersion (CD) or group velocity dispersion highlighted in section 1.1 [7]. The CD happens due to different propagation velocities of the wavelength components of the signal spectrum from the refractive index dependency on the used wavelength. In other words, an optical pulse will spread in the time domain during propagation. Fortunately, this impairment is a static impairment, where it can be compensated knowing the dispersion parameter and the dispersion slope that characterize the optical fiber through which the propagation occurs.
- Polarization mode dispersion (PMD) and polarization rotation [7]. The PMD happens due to the imperfection of the cylindrical shape of the core fiber, i.e., the axes of the core fiber are not symmetric. This imperfection is known as fiber birefringence. The asymmetry in the fiber axes results in slight deviation in the velocity of the polarization modes on those axes leading to PMD effect on the propagating signal. In addition, the fiber birefringence axes randomly rotate along the fiber. The overall effect of this birefringence fluctuation is a polarization rotation of the input signal to the fiber such that the output signal has a random state of polarization that differ from the state of polarization of the input signal.
- Nonlinear impairments [7]. If the signal is propagating in a single carrier transmission, it suffers from a nonlinear effect, called self phase modulation (SPM). If there is a WDM transmission, the signal on a specific channel is affected by the neighboring channels by another sort of nonlinear effect, called cross phase modulation (XPM). Those effects in general depend on the power profile and the power level of the propagating signals. There are other nonlinear effects on the optical signal depending on the scenario used in the system, e.g., four wave mixing (FWM), stimulated Raman scattering (SRS), and stimulated Brillouin scattering (SBS). More details about the nonlinear impairments can be found in [7].

Next, we review the basic DSP stack used in this thesis for coherent system. Indeed, all DSP tasks can be done at the receiver side. However, introducing some of this DSP at the transmitter side enhances the system performance. For example, performing pulse shaping at the transmitter side helps in mitigating the inter-symbol interference (ISI) and enhances the system performance. A raised cosine (RC) pulse shaping provides a band-limited spectrum and zero ISI [65]. Typically, the RC pulse shaping filter is split into two matched root-raised cosine (RRC)

filters at the Tx-DSP and Rx-DSP to maximize the signal-to-noise-ratio (SNR) assuming an additive white Gaussian noise (AWGN) channel. In addition to pulse shaping, the Tx-DSP may include equalization of the frequency response of the DAC, RF amplifier and IQM, and non-linear compensation of the transfer function of the Mach-Zehnder I-Q modulator shown in Fig. 1.4(a). The Rx-DSP compensates for imperfections in the Rx front-end, chromatic dispersion associated with the optical signal propagation in the fiber. In addition, Rx-DSP performs matched filtering if pulse shaping is done at the Tx-DSP, time recovery, frequency offset removal, phase noise mitigation and polarization de-multiplexing as shown in Fig. 1.4(b). A more detailed discussion for the mentioned DSP in addition to the used algorithms can be found in [66-76].

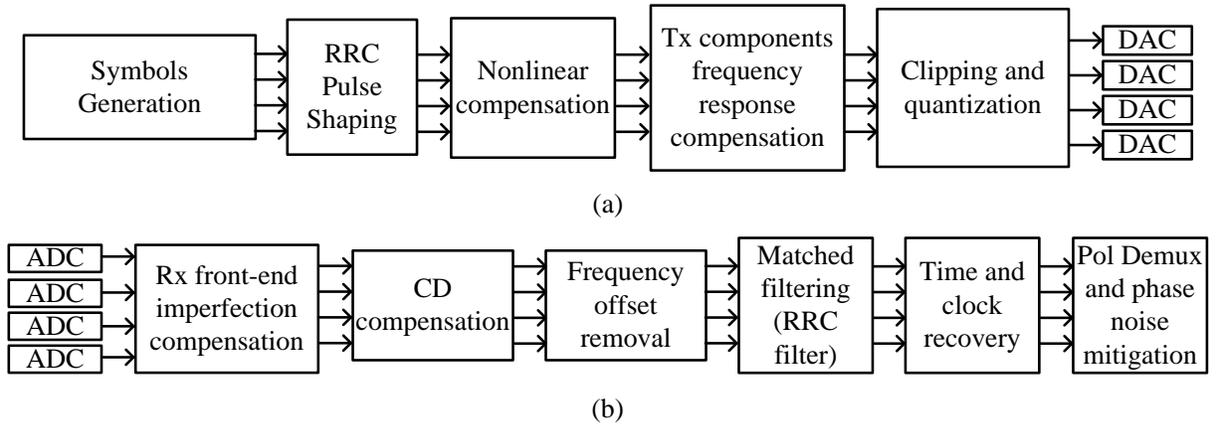


Fig. 1.4 Basic DSP stack: (a) Tx-DSP, and (b) Rx-DSP.

1.2.3.2 Stokes Vector Direct Detection (SV-DD) system

In optical communications, there are two orthogonal state of polarizations (SOPs) that can carry two independent signals. The transmitted field of the optical signal can be represented either using Jones vector $E^t = [E_x^t \ E_y^t]^T$ or Stokes vector $S^t = [S_0^t \ S_1^t \ S_2^t \ S_3^t]^T$, where the four Stokes parameters are related to Jones vector by $S_0^t = |E_x^t|^2 + |E_y^t|^2$, $S_1^t = |E_x^t|^2 - |E_y^t|^2$, $S_2^t = 2\text{Re}\{E_x^t E_y^{t*}\}$, $S_3^t = 2\text{Im}\{E_x^t E_y^{t*}\}$, and T is the vector transpose. In this subsection, we are focusing on representing the signal in Stokes space where all Stokes parameters are real and can be directly detected using photodetectors (PDs). SV-DD system has emerged as a possible solution to increase the data rate on single carrier for short reach with respect to IM/DD systems and to simplify the complexity with respect to coherent detection [58, 77-83]. In short reach systems, it allows for polarization multiplexing and de-multiplexing for IM/DD system using

multiple-input-multiple-output (MIMO) DSP [83]. In another case, the second polarization may not be modulated such that it is used as a carrier for signal beating at the receiver [81]. However, the used receiver optical front-end is still the same for detecting the Stokes vector representation of the received signal and using the DSP the transmitted data is retrieved.

In [82], possible transmitter configurations to transmit data that can be retrieved by SV-DD have been presented along with possible receiver optical front ends for SV-DD. The IM/DD is a unidimensional (1D) modulation where the intensity of a CW laser power is either modulated directly or using external modulator at the transmitter, and a PD is used at the receiver to detect the modulated power. The Stokes vector direct detection system using one hybrid and 3 BPDs at the receiver extends the modulation dimensions to 2D and 3D on a single carrier. The 2D modulation can be either IM on the two orthogonal polarizations [83] or complex modulation on one polarization using IQM and sending a reference carrier on the other polarization [81]. From the Stokes vector equations, the two IM signals in the first 2D modulation configuration can be retrieved by detecting S_0^t and S_1^t . However, after optical signal propagation in the fiber, the received Stokes is rotated by a rotation matrix R , where $S^r = RS^t$. Thus, MIMO DSP needs to have access to the other two Stokes parameters to de-rotate the received Stokes and retrieve the two IM signals. In the second 2D modulation configuration can be retrieved by detecting S_2^t and S_3^t . In this configuration, after de-rotation S_2 and S_3 represent the electric field of the modulated polarization which allows for compensating the chromatic dispersion. It is worth noting that $S_0^t = S_0^r = \sqrt{(S_1^2 + S_2^2 + S_3^2)}$, where S_0 is representing the total power of the signal. Thus after normalization of the transmitted and received signal powers the first row in the rotation matrix R is $[1 \ 0 \ 0 \ 0]$. Thus, the maximum degree of freedom without any additional hardware is 3D which can be achieved by several configurations shown in [82]. An attempt to achieve 4D modulation using differential phase between two successive symbols in Stokes space is demonstrated in [84]. This requires additional delay and an additional $2 \times 4 \ 90^\circ$ hybrid. In this thesis, we focus on using the second configuration of 2D modulation illustrated in Fig. 1.5 with the corresponding receiver configuration.

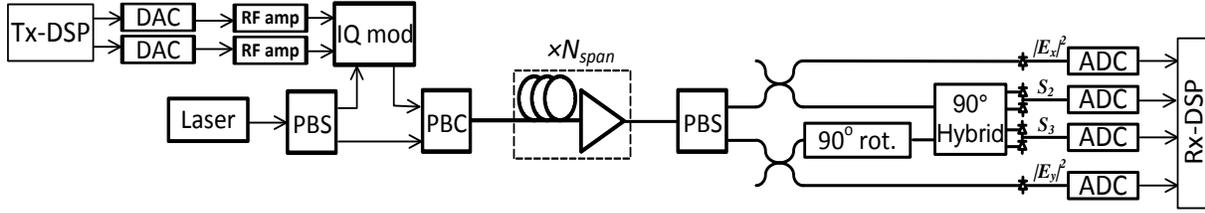


Fig. 1.5 Schematic for 2D configuration for SV-DD system.

1.3 Problem statement and thesis objectives

In this thesis, we aim at investigating and proposing possible optical solutions for DCIs operating in intra- and inter-datacenter networking. In the first part of this thesis, we address the intra-datacenter networking regarding two important aspects; (i) the optical switching implementation in the core layer providing high-speed inter-pod communication, and (ii) the Ethernet frames transmissions speed evolution per wavelength with small form factor pluggable solutions.

As aforementioned in section 1.1, optical switching is a potential solution to have switching mechanism that is transparent to the data rate at the input and output links and at the same time, reducing the power consumption significantly by reducing the number of required transceivers in the core layer. However, optical switching is immature in terms of buffering to resolve the contention problem, the required overhead associated with extracting the header to be processed by the control unit, and buffering the data packet while the header is processed. Thus, we aim at proposing possible network architecture for appropriate implementation of optical core switching. More specifically, instead of focusing on how to resolve those problems in the core layer, our objective is to utilize the network as a whole providing a complete solution for optical switching implementation. In other words, we focus on the system level design and associated control plane requirements, and on how the physical layer optical components serve within this proposed network architecture.

Next, the evolution of the IP traffic in the intra-datacenter networking requires an equivalent revolution for the capacity per link. The proposed solutions for IM/DD system are limited in terms of spectral efficiency and the implementation of the IM/DD system requires operating at low symbol rates on multiple wavelengths or fiber lanes to achieve with the required data rate at low BER to minimize the FEC overhead. In this thesis, we target enhancing the spectral efficiency in the intra-datacenter networking by utilizing the four available dimensions for modulation in the optical communications. In parallel, we investigate the use of an integrated

solution at the transmitter; namely InP DP-IQM providing a potential compact transceiver for the proposed system.

In the second part of this thesis, we address the challenges in inter-datacenter networking. As highlighted in subsection 1.2.3.1, despite some of the common challenges between the intra-datacenter networking and inter-datacenter networking, the main challenges of reach and capacity lead to the adoption of the optical coherent system. As a first step in addressing one of the common challenges of having cost efficient integrated pluggable solution, we investigate the viability of the use of InP devices to achieve high-speed transmissions with integrated small form factor solutions. In addition, we address the limited bandwidth of the cascaded electro-optic components, which is considered as the bottleneck of high-speed transmissions. The main objective is to use the evolution of the DAC and ADC technology to provide high-speed transmission enabled by the cumulative algorithms used in the Tx-DSP and Rx-DSP to handle the transceiver impairments. As a second step, we investigate a 2D-SV-DD system as an alternate solution for inter-datacenter applications with reduced complexity at the expense of halving the capacity compared to coherent systems. The main objective is to extend the capacity and reach of this alternative solution compared to the available demonstrations in the literature for the same system.

1.4 Original contribution

In view of the discussed objectives, the original contribution of this thesis can be summarized as follows:

- We propose a contention resolution strategy in optical burst switched intra-datacenter networking with centralized scheduling. The main idea is to use the electronic burst assembly buffer at the transmitter for burst construction and for contention resolution. In addition, we propose an AWGR-based switching mechanism, where tunable lasers are used at the transmitter side for both switching and optical transmission. Thus, we resolve the required complexity at the core switch by moving it to the source transmitter with the aid of the centralized scheduling.
- We propose a self-homodyne system for intra-datacenter networking. The proposed system enhances the spectral efficiency of the modulated signal with respect to the IM/DD systems and uses DSP of reduced complexity from the conventional coherent

system. The concept of the proposed system is to send the coherent modulated signal and a tone originating from the same laser over the full-duplex fiber using circulators for coherent detection at the receiver.

- We demonstrate experimentally 400G and 600G transmissions using an InP DP-IQM on a single wavelength. The demonstrations are done using two generations of high speed DACs. The first set of results show that a 56 Gbaud PDM-16QAM modulated signal can be transmitted over 500 km at a BER below the hard decision forward error correcting threshold of 3.8×10^{-3} . The second set of results report a record breaking transmission of 770 Gb/s (77 Gbaud PDM-32QAM) over 320 km of standard single mode fiber (SSMF), which is suitable for low cost DCIs in inter-datacenter networking.
- We propose to do the chromatic dispersion compensation at the Tx-DSP (CD pre-compensation) in 2D-SV-DD system, where a pilot carrier is polarization-multiplexed with a single polarization complex modulated signal. Generally, the 2D-SV-DD system has reduced Rx-DSP complexity compared to coherent detection. The CD pre-compensation allows the merge of two DSP steps required at the receiver in the MIMO DSP block, namely; (i) performing the polarization de-rotation and (ii) mitigating any residual ISI. Next, we extend our study by comparing, aided with mathematical analysis, two suitable methods for CD compensation, namely CD pre- and post-compensation in the above-mentioned 2D-SV-DD system. Results show that the CD pre-compensation can be used for any transmission distance, where we are able to report the highest throughput-times-distance product achieved in Stokes transmission systems. For example, 84 Gbaud QPSK and 64 Gbaud QPSK transmissions reached 3,520 km and 4,800 km below the BER of 2×10^{-2} achieving 591,360 Gb/s.km and 614,400 Gb/s.km throughput-times-distance products, respectively. However, the use of CD pre-compensation degrades the output signal quality from the DAC, especially when the symbol rate is closer to the DAC sampling rate. For example, when an 84 Gbaud 16QAM signal is transmitted using a DAC running at sampling rate of 84 GSps and detected using a 2D-SV-DD receiver, BER after 640 km propagation is 1.67×10^{-2} and $\sim 2 \times 10^{-2}$ in case of CD post- and pre-compensation, respectively.

1.5 Thesis organization

After the introductory chapter, the rest of this thesis is organized as follows:

- In **Chapter 2**, we present the first contribution of this thesis for the implementation of optical burst switching in intra-datacenters with the proposed contention resolution strategy. We present the proposed network architecture with centralized scheduling in two cases using active fast switching fabrics and using the AWGR-based switching mechanisms. We conduct simulations using OPNET to prove the effectiveness of the proposed contention resolution strategy. In addition, we investigate the effect of self-similarity on the network performance and the scalability of the network resources in terms of number of switches and wavelengths required in case of not using the proposed contention resolution strategy to achieve the same throughput using the proposed contention resolution strategy with limited network resources.
- **Chapter 3** represents the second contribution in this thesis proposing the self-homodyne system for intra-datacenter networking. The principle of the proposed system is discussed and a summary of the system advantages is presented. We experimentally demonstrated the proof of concept of the proposed system using emulation of the bi-directional propagation of the tone and the modulated signal. We report, to the best of our knowledge, record breaking transmissions of 530 Gb/s, 448 Gb/s, and 320 Gb/s over 500 m, 2 km, and 10 km of SMF fibers, respectively, below KP4 threshold using integrated InP modulator.
- In **Chapter 4**, we present the high-speed optical coherent transmissions enabled by DSP using InP DP-IQM for inter-datacenter networking. The experimental demonstrations show the viability of using the InP DP-IQM in a low cost DCIs solution for high-speed inter-datacenter communication. Digital pre-emphasis optimization is one of the most important DSP blocks that enhances the system performance. For the 84 Gbaud signals, slight IQ imbalance and IQ skew can affect system performance. Since the perfect matching cannot be achieved especially when using discrete components, we used a 4×4 real valued MIMO to compensate for any residual IQ skew and/or imbalance.
- In **Chapter 5**, we present the last contribution in this thesis with a potential alternative solution for optical coherent system in inter-datacenter communications; namely 2D-SV-DD system. We propose the use of CD pre-compensation at the transmitter allowing for

extended reach compared to recent demonstrations in the literature for similar 2D-SV-DD system. In addition, we investigate the impact of CD compensation methods on the system performance.

- Finally, we conclude the thesis in **Chapter 6** and introduce potential avenues for future work related to the covered topics in this thesis.

Chapter 2

Optical Burst Switching in Intra-Datacenter Networking

2.1 Overview

A large datacenter consists of a number of building blocks (modular containers) called pods (performance-optimized datacenters). Each pod contains around one thousand servers that are optimally connected to maximize connectivity, and minimize cost, cooling and power consumption [18, 85]. One of the main challenges in intra-datacenter networking is providing a full bisection bandwidth in core layer switching to achieve inter-pod communications. Many approaches aimed at handling this challenge by designing scalable topologies using electronic switching as in [29, 31, 86]. Unfortunately, providing full bisection bandwidth for all-to-all communication using electronic switching is difficult due to limited bandwidth, wiring complexity, and power consumption.

The use of optical switching is necessary not only to meet current challenges, but also to provide reliable solution for next generation networks in datacenters. Optical switching can meet the ever-increasing bandwidth demand, provide data rate transparency [87], and reduce cost, and power consumption. The reduction in cost and power consumption is attributed to the absence of optical transceivers required for O/E and E/O conversion at the core layer electronic switches. One of the proposed solutions for optical switching implementation inside datacenters is to use optical switching side-by-side with electronic switching [18, 30, 88]. The main idea is to use optical or electronic switching whichever is appropriate according to traffic pattern. For example in [18], a hybrid electrical/optical architecture for datacenters has been proposed where they have used optical circuit switching (OCS) in the inter-pod switching. The choice between optical and electronic switching in that case depends on the aggregated traffic characteristics. However, these proposals are still using electronic switching, and need traffic prediction to choose between OCS and electronic switching.

Recently, all-optical switching implementation in intra-datacenter communications and high performance computing networks has become one of the cutting-edge technologies that concerns

researchers in Datacom [15, 22-26, 29, 32, 33, 89]. In [22], a single AWGR equipped with TWCs at its input ports has been used for switching. Optical label switching has been used, where the optical label is extracted and forwarded to a central control unit while the data packet is optically delayed in a FDL until the label is processed; scheduling is done and the TWC is configured by the control unit. If contention occurs, packets are buffered in a SDRAM after O/E conversion. When the contention is over, packets are retransmitted. In [23], an AWGR equipped with TWCs at its input and output ports has been used as a single switch in a multistage switching system. FDLs are used instead of SDRAM for contention resolution. EDFA are used to make the signal power level large enough as an input to the TWC. In [32], labeled optical burst switching with home circuits (LOBS-HC) has been used where the burst is constructed according to minimal-size based (burst-length-based) algorithm. The system topology used is different from the conventional system where core switches are interconnected using a hypercube connection topology. In [33], a similar approach to what we proposed in [90] has been used without handling the physical layer concerns. The use of centralized control along with the burst assembly buffers to resolve the contention is adopted. A horizon scheduling is used based on a synchronous system to schedule the burst transmission based on prospective connection releases. The centralized scheduler assigns the wavelength and the start time for burst transmission for the transmitting pod.

In the implementation of buffering in conventional switching, only first-in-first-out (FIFO) buffers are used at input ports which causes a problem called head of line (HOL) blocking problem [91]. A well-known approach to solve this problem is to use virtual output queues (VOQs), where there is a queue for each output port at each input port so that the packet at the head of the queue does not block the way to serve other packets destined to other output ports behind it. Thus, scheduling algorithms are developed to choose the packet to be served at each time slot. Different algorithms are reported in literature such as maximum size matching algorithm, parallel iterative matching algorithm, round robin matching algorithm, and iSLIP algorithm [92]. These algorithms are suitable for achieving high throughput with uniform traffic. There is another group of algorithms called maximum weight matching algorithms where each queue at the input is given an updated weight for service priority. Two algorithms that are special cases of this type are longest queue first (LQF) and oldest cell first (OCF) [93]. These two

algorithms achieve 100% throughput, however, only OCF is suitable for non-uniform traffic because LQF can cause starvation for some queues in some cases.

In this chapter, we focus on switching between pods in intra-datacenter networking. We use a centralized signaling protocol in the control plane taking advantage of the co-locality of the pods in the datacenter [11] instead of the one-way reservation protocols [10, 12, 94]. The one-way reservation protocols have been used in most demonstrations of OBS in wide area networks (WANs) taking into consideration the impossible realization of centralized control and the two-way signaling protocols [10]. OBS is adopted as the switching technique, where the burst is the basic data unit that is constructed by aggregating a group of packets that have the same destination using a certain assembly algorithm. We use hybrid burst assembly algorithm to construct the data bursts since it provides the least latency due to assembly compared with other algorithms with the same parameters [10]. Hybrid burst assembly algorithm is simple and suitable for datacenters where all pods are co-located in the same area. On the other hand, other complex and advanced burst assembly algorithms were mainly developed due to the geographical separation between the ingress nodes and the core nodes in conventional OBS WAN networks in literature [10, 95]. We propose a contention resolution strategy where we use the assembly buffers not only for the burst construction but also for contention resolution where the bursts are kept in the assembly buffers until the centralized scheduler sends a response for the corresponding burst transmission request [90]. The main idea is to benefit from one of the main concepts of OBS, which is not to go to the optical domain except when connection is guaranteed. The concept of this contention resolution strategy is primarily illustrated in a network assuming the availability of sufficient optical space switches to build the required switch banks in the proposed system. We will call this approach from now on “active approach” to address the specific assumptions in the simulation conducted for this phase and describing the associated network architecture [90]. Next, we propose a “passive approach” using AWGR-based mechanism where single or multiple AWGRs can be used as passive optical spine switches in the core layer. The active part where the configuration takes place is moved from core switches to the pod transmitter side where tunable lasers (TLs) are used in the E/O conversion at the pod-to-core interface. This keeps the optical core switches passive by omitting the use of TWCs enabled by replacing the fixed wavelength lasers at the transmitter in the “active approach”, which are already active elements, with TLs that can be tuned according to the destination. It is worthy to

say that the existence of pods and switches in the same place enables complexity transfer from the optical switch to the transmitter and the centralized scheduler. In other words, the overhead required to extract the control header from the optical signal is omitted and the required delay for the data portion of the optical signal is not needed by having a centralized scheduler with dedicated links. Finally, employing the electronic assembly buffers replaces the role of the input queues in the conventional switching for resolving the contention. We use an oldest burst first (OBF) algorithm in case of contention, which is a modified version of the OCF algorithm.

Simulation is done using OPNET to evaluate the performance of the proposed network. In the “active approach”, we assume that the traffic introduced by each pod has the same parameters. In addition, we consider that Ethernet packets arrive to the pod interface, such that packet inter-arrival times follow a lognormal distribution [96]. In the “passive approach”, two cases are investigated for different traffic loads: 1) the inter-arrival time between packet arrivals to the pod interface follows an exponential distribution implying Poisson arriving traffic as generally assumed in most standard telecommunication networks, and 2) the packet inter-arrival time follows lognormal distribution, which agrees with the measurements in [96] that characterize datacenter traffic more accurately compared to exponentially distributed inter-arrivals. We investigate the impact on the network performance due to self-similarity when the traffic is generated for each source pod using lognormal distributed inter-arrival times. In addition, we investigate the effect of increasing the electronic buffer size on network performance. The increase of buffer size improves throughput at the expense of larger latency, which may reach ~1 msec to achieve 100% throughput.

The remainder of this chapter is organized as follows: In section 2.2, we propose the network architecture used in both aforementioned approaches. Section 2.3 is dedicated for describing the used burst assembly algorithm. Next, we discuss the self-similarity feature of traffic based on lognormal distribution in section 2.4. Section 2.5 describes the control plane including resource allocation and the modifications in the OBF algorithm. Finally, Section 2.6 presents the simulation results, and we finally conclude the chapter in section 2.7.

2.2 Network architecture approaches for OBS in intra-datacenter networking

In this section, we show two different network architecture approaches to apply the proposed contention resolution strategy. The active approach discussed in subsection 2.2.1 assumes the use of fast fabric switch like in [97] as the basic switching element in the core layer. Next, we present the network architecture of the passive approach supporting the proposed AWGR-based switching mechanism in subsection 2.2.2. The two approaches adopt the same idea of using the electronic assembly buffers to resolve the contention. However, in the passive approach, we further simplify the requirements in the core layer switch where no communication is required between the centralized scheduler and the core layer to configure the switches. The tunable lasers are used at the transmitter for burst transmission, and are tuned at the appropriate wavelength for switching based on the scheduler response.

2.2.1 Active approach for OBS network architecture in intra-datacenter networking

A conventional network topology used in datacenters is shown in Fig. 2.1(a). In our proposed approach, the network topology in Fig. 2.1(a) is maintained except for minor modifications in the architecture of the pod-to-core interface and in the core switching as outlined in Fig. 2.1(b) and Fig. 2.1(c), respectively. Essentially, pod switches in the datacenter serve as ingress and egress nodes in the OBS network, whereas the core layer switches serve as core nodes in the OBS network. For the pod-to-core interface, a set of $(N-1)$ electronic assembly buffers, corresponding to the number of possible destinations, are needed at each pod switch, where N is the number of pods. Packets arriving at the pod switches are buffered in one of the available buffers according to their destination pod for burst construction using hybrid burst assembly algorithm whose operation will be described later in section 2.3. In addition, an electronic switch controlled by the scheduler is used between the $(N-1)$ buffers and the $(w \times S)$ optical transmitters (inside the E/O). The E/O conversion process occurs after the burst is constructed in the electrical domain as shown in Fig. 2.1(b). Fixed wavelength lasers can be used for the E/O blocks in the active approach where either direct or external modulation can be employed. On the other hand, the core layer consists of S switch banks, each containing w $(N \times N)$ fast switch fabrics [97], N wavelength demultiplexers at the input of the switch bank, and N wavelength multiplexers at the

output of the switch bank. The switch bank has N input ports and N output ports corresponding to the N pods. Each input/output port is connected to a wavelength demultiplexer/multiplexer. Each switch fabric in the switch bank is used to make a connection at a specific wavelength between input and output ports. When the optical signal arrives at the input port, it is demultiplexed to its w wavelengths. Then, each wavelength is directed to its corresponding switch fabric that is already pre-configured by the central scheduler according to the destination pod. After the switching process occurs, the optical signal is directed to the multiplexer dedicated to the destination pod. For example, when a burst is to be transmitted from pod n to pod m where $n, m \in \{1, 2, \dots, N\}$ and scheduled on wavelength k where $k \in \{1, 2, \dots, w\}$, the optical signal from pod n is demultiplexed at input link i_n . The wavelength λ_{ink} is directed to the k^{th} switch in the switch bank, which is configured by the scheduler to switch the signal to the destination pod m on λ_{omk} . The wavelength λ_{omk} is directed to be then multiplexed at the output link o_m .

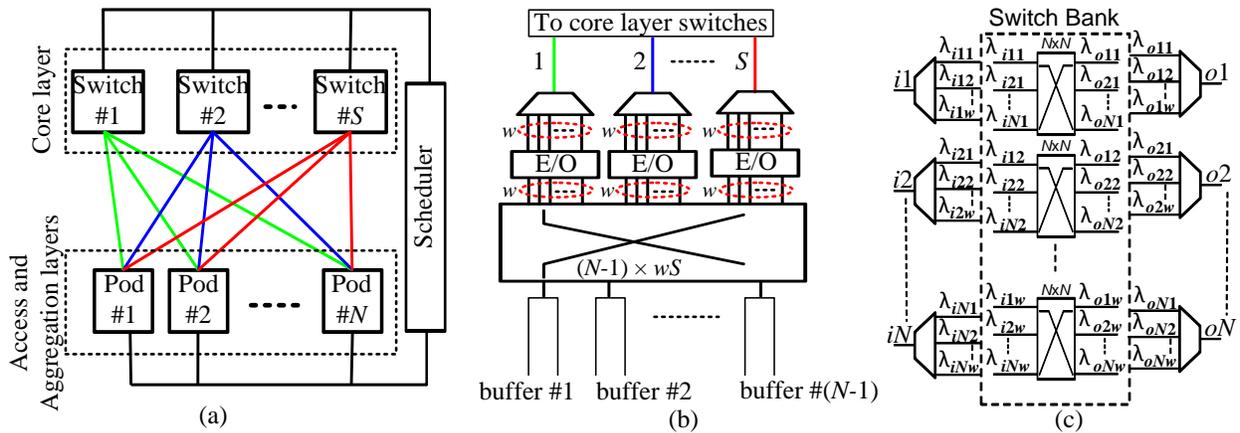


Fig. 2.1 Active approach for OBS network architecture: (a) Main intra-datacenter network topology with centralized scheduling; (b) pod interface to core layer switching; (c) core switch bank based on active fast fabric switches.

2.2.2 Passive approach for OBS network architecture in intra-datacenter networking using AWGR-based switching mechanism

In the active approach introduced in subsection 2.2.1, fabric switches in switch bank aided with multiplexers/demultiplexers deployed at its input/output ports are used to connect between source pods and destination pods and this is done at each wavelength to scale up the bandwidth, i.e., each switch bank comprises w switch fabrics. In the active approach, the task of the switch bank is the ability to have multiple simultaneous connections on multiple wavelengths between

the input and output ports of the switch bank while using space switch fabrics as the basic switching units. In our proposed AWGR-based switching mechanism denoted hereafter by “passive approach,” we use the AWGR to perform this task benefiting from the AWGR properties described in subsection 1.2.1.2. Within the same FSR, switching can be performed from any input to any output of the AWGR, which is the task of a fabric switch in the switch bank. The repetition of the AWGR wavelength-based routing property every FSR allows for using several FSRs to multiplex several simultaneous connections at the input and output ports of the AWGR. Thus, the number of multiple FSR periods used in the passive approach replaces the number of wavelengths in the active approach.

In [27], an AWGR switching mechanism is used such that the number of ports of the core switch can be increased while using AWGR with a smaller number of ports by using different FSRs in parallel. In our proposed mechanism, we use the FSR to scale up the bandwidth that can be used by each source pod to communicate with different pods through different FSRs. Thus, when an FSR is used to connect a source pod to a destination pod, this FSR cannot be used by either of them in another connection until a connection release. However, another FSR can be used in another connection with other pods. In summary, in our approach, the single AWGR acts as several space switches at different FSRs, while in [27] different FSRs are used to increase the number of physical ports at the core switch.

Fig. 2.2 shows the proposed network architecture for the passive approach. Essentially, pod switches in the datacenter serve as ingress and egress nodes in the OBS network, whereas the core layer switches (AWGRs) serve as core nodes in the OBS network as shown in Fig. 2.2(a). Thus, a set of $(N-1)$ electronic buffers, shown in Fig. 2.2(b), is needed for burst construction at each pod switch where N is the number of pods. In addition, an electronic switch controlled by the scheduler is used between the $(N-1)$ buffers and the $(k \times S)$ optical transmitters (inside the E/O), where k is the number of FSRs and S is the number of AWGRs. Packets arriving at pod switches are buffered into one of the available buffers according to their destination. For each burst transmission, the scheduler configures the electronic switch, shown in Fig. 2.2(b), according to the assigned FSR and AWGR (core switch) followed by E/O conversion. The E/O block described in Fig. 2.2(c) is a group of tunable transmitters with TLs. Each TL is dedicated to a certain FSR and is tuned to feed the modulator with a wavelength within the FSR according to the destination. This allows the transmitter to make several connections at the same time

according to the number of used TLs (used FSRs). After modulation, the signals are multiplexed into one fiber directed to the assigned AWGR. At the receiver side, the optical signal is demultiplexed to the k used FSRs, i.e., it behaves as a filter for the FSR at each branch. Then, O/E conversion is done only at the destination, which saves power in the core layer since the data plane operates optically. It is worthy to say that the network capacity or overall throughput can be scaled up in several ways: i) increasing the number of used FSRs, ii) increasing the number of AWGRs, or finally iii) increasing the data rate of the transmitted signal.

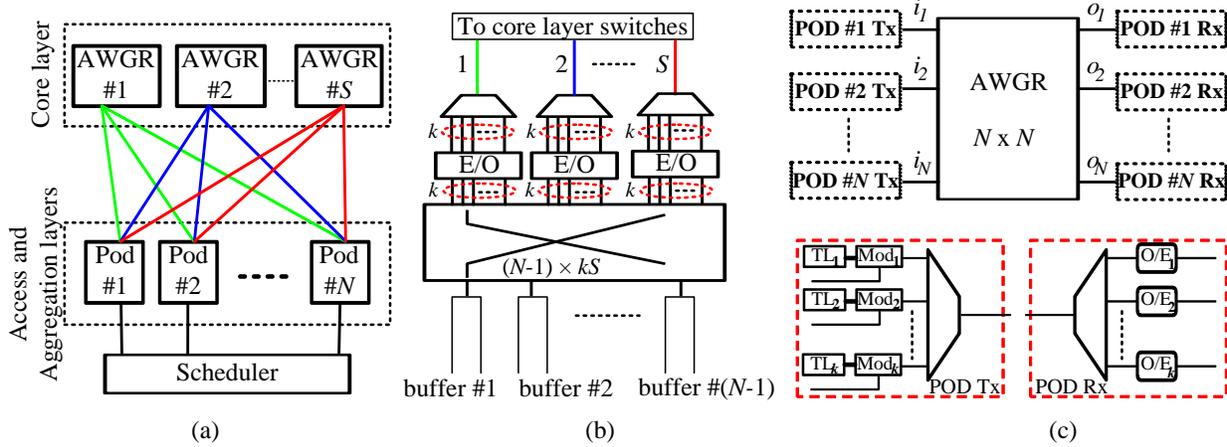


Fig. 2.2 Passive approach for OBS network architecture: (a) Main intra-datacenter network topology with centralized scheduling (not connected to core layer); (b) pod interface to core layer switching; (c) AWGR-based switching mechanism with E/O and O/E block description.

2.3 Hybrid burst assembly algorithm

The burst comprises a number of packets in OBS networks. Upon construction, the data burst constitutes the basic data unit to be switched optically by core network switches. The construction of the data burst follows a certain assembly algorithm according to which data packets are assembled into the data burst based on their destination. There are several burst assembly algorithms; time-based, burst length-based, and time/length-based (hybrid) assembly algorithms [8, 10]. In the time-based assembly algorithm, upon the first arrival of the packet to the assembly buffer, a timer is started and the arriving packets are assembled. The burst is constructed, when the timer reaches the time threshold (T_{th}). In the burst length-based assembly algorithm, a byte counter is started instead of the timer and when the burst size reaches the burst length threshold (B_{th}), the burst is constructed. In the hybrid algorithm, which is adopted in this work, a timer is started as packets are assembled. If the burst size reaches a burst length

threshold (B_{th}) before the time is up, the burst is constructed. Otherwise, the burst is constructed when the timer reaches the time threshold (T_{th}).

The use of the hybrid algorithm gives the optimum latency compared to the other two assembly algorithms since it uses the earlier of the two thresholds. As shown in Fig. 2.3, which is a redraw taken from [8, 10], when the length-based assembly algorithm is used in high traffic scenario, the time at which the burst is constructed is t_1 , which is smaller than T_{th} at which the burst is constructed in case of the time-based assembly algorithm. However, in the light traffic scenario, the burst is constructed by using the time-based assembly algorithm at T_{th} , which occurs earlier than its construction by length-based assembly algorithm at time t_2 . On the other hand, a hybrid assembly algorithm picks the earlier of the two thresholds in either traffic case, i.e., it constructs the burst at t_1 and T_{th} for the heavy and low traffic scenarios, respectively. We can conclude that the choice of the burst assembly algorithm and setting its thresholds has a great impact on the overall network performance (e.g., latency). In addition, it also impacts the shape and characteristics of the burst traffic generated by the ingress node which will then arrive at the core switches. This traffic shaping property of the burst assembly process is especially useful when the packet arrival process to the assembly buffers has self-similar characteristics, which can be smoothed via the assembly process generating a smoother burst traffic that is less self-similar as will be pointed out later in the upcoming subsection.

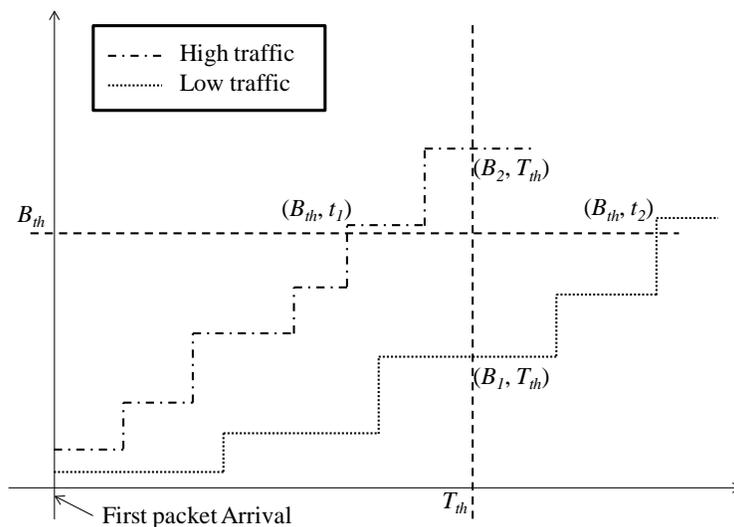


Fig. 2.3 Hybrid Burst Assembly Algorithm.

2.4 Self-similarity and burst assembly

When traffic is said to be non-self-similar, this means that packet inter-arrival times follow a memoryless distribution like exponential distribution. In non-self-similar traffic, when time scale gets larger, the overall traffic behavior tends to give only the mean value, i.e., the randomness can be removed by averaging over longer time scales. On the other hand, when traffic is said to exhibit self-similarity, this means that traffic at different time scales has similar behavior, where correlation exists between different time scales. Thus, self-similar traffic exhibits randomness that does not get smoothed by averaging over longer time scales, i.e., the variance of the sample space is relatively high. This can be seen in the probability density function of the so-called heavy-tailed distributions. Leland *et al.* explained the self-similar characteristics of Ethernet traffic in [98].

Benson *et al.* have made measurements on traffic traces over several datacenter networks [96]. They deduced that the datacenter traffic has ON/OFF periods that follow lognormal distribution. Here the ON period means that the entity produces traffic where data units are generated with random inter-arrival times. This random inter-arrival time follows lognormal distribution. On the other hand, the OFF period means an idle period where the entity does not produce any traffic for a period of time. In [99], Fernandes *et al.* tested several distributions that results in self-similar traffic among which is the lognormal distribution. They reported that every distribution has the so-called shape parameter, which controls the degree of self-similarity of the resulting generated traffic. For the lognormal distribution, there is a relation between the probability distribution function (PDF) and the associated normal distribution function. The PDF of the lognormal distribution $P(x)$ can be represented as follows:

$$P(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}}$$

where x represents the generated random variable, μ and σ are the mean and the standard deviation of the associated normal distribution. The standard deviation (σ) is the shape parameter that controls the degree of the self-similarity of traffic generated using the lognormal distribution. The shape parameter σ is related to the mean and variance of the lognormal distribution according to the following equation:

$$\sigma = \sqrt{\ln\left(\frac{v}{m^2} + 1\right)}$$

where v and m are the variance and the mean, respectively, of the generated samples following the lognormal distribution.

Fig. 2.4 shows the lognormal PDF at different values of σ while the mean is kept the same at 1.6487 (which is the value of m when $\mu = 0$, and $\sigma = 1$). The larger the value of σ , the greater the probability of occurrence of smaller values of x below the sample mean (m). For example, in case of $\sigma = 1.5$, 75% of the sample values are below the sample mean (m). Thus, when the packet inter-arrival times follow a lognormal distributed sample with a shape parameter σ , the larger the value of σ , the higher the probability of having inter-arrival times smaller than the mean inter-arrival time. In addition, the variance of the lognormal distribution will be higher such that a higher sample value compensate for the concentration of the sample values below the sample mean. This explains why the performance is not predictable based solely on the mean inter-arrival time for self-similar traffic, i.e., lognormal distributed inter-arrivals, and it is generally worse than the performance in case of non-self-similar traffic, i.e., exponential distributed inter-arrivals.

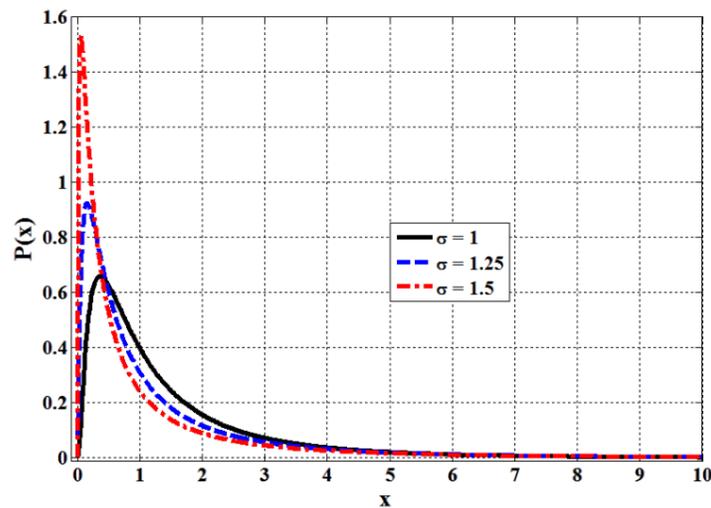


Fig. 2.4 Probability Distribution function for lognormal distribution at different shape parameters.

Although the shape parameter is important in self-similar traffic generation, it is not a direct metric to quantify the degree of self-similarity of the traffic. Another parameter called Hurst parameter (H) is used to measure the degree of self-similarity of the random process under test [98]. The random process (here the generated traffic) is said to be self-similar when $0.5 < H <$

1. The least self-similarity happens when $H \approx 0.5$ and the maximum self-similarity occurs when $H \approx 1$.

To conclude this subsection, burst assembly is a way to regulate the inter-arrival time distribution, in other words, shape the traffic, where packets are aggregated in one data unit called burst. Clearly, the burst assembly delays the packet transmission due to the underlying assembly latency. The resulting inter-arrival times between bursts after traffic shaping due to assembly is generally larger than the packet inter-arrival time of the input traffic, i.e., the burst arrival rate at a core switch is less than the packet arrival rate at the assembly (edge) node. At this point, it seems the traffic load is decreased however although the burst arrival rate is typically less than packet arrival rate, the mean service time of a burst is longer than the mean service time of a packet because of the difference in their respective durations. Thus, the offered load will remain the same after burst assembly, however, the mean value of inter-arrival time will be more representative of the network performance due to the reduced self-similarity after the assembly because of the reduction of occurrence of values less than the sample mean. Intuitively, the larger the assembly thresholds, the smoother, i.e. less self-similar, the resulting burst traffic at the expense of additional latency due to burst assembly.

2.5 Control plane layer for OBS in intra-datacenter networking

As previously highlighted, a centralized signaling control is adopted taking advantage of the locality of pods within the datacenter. The centralized scheduler has a global knowledge of the status of the core layer switch banks/AWGRs in the datacenter. Also, it keeps track of which wavelengths/FSRs are used at the input and output links of those switch banks/AWGRs. Here, the centralized scheduler has three tasks: 1) the switch bank/AWGR assignment which is done using a round-robin algorithm, 2) the wavelength/FSR allocation which follows a first-fit algorithm at that switch bank/AWGR for this connection, and 3) the scheduling related to handling connection requests. The third task is concerned with how the scheduler deals with multiple connection requests queued in the control layer. We employ a first-come-first-served (FCFS) scheduling algorithm to handle new requests, while oldest-burst-first (OBF) scheduling algorithm is used for handling contending requests, which were queued, upon connection releases.

2.5.1 Resource allocation for new connection requests

After burst construction, the source pod sends a control packet to the scheduler requesting burst transmission to a destination pod. Initially, the scheduler checks the request queue related to this source/destination pair. If this queue is not empty, this means that the connection between this pair cannot be setup and this new request will be buffered. For new request arrivals that are not buffered in the queues dedicated for contention, they are served using FCFS scheduling algorithm where priority is given to the request that first arrived to the centralized scheduler.

2.5.1.1 Switch bank/AWGR allocation

When a burst is constructed, the source pod sends a connection request to the central scheduler. This request contains three fields: source field, destination field, and burst length field. If $S > 1$, there is a need for a certain algorithm to be followed by the scheduler to assign a certain switch bank/AWGR to the new connection request. We adopt a round-robin algorithm for the switch bank/AWGR assignment, where the priority is given in the next assignment to the switch bank/AWGR that follows the last assigned switch bank/AWGR in the previous connection. This algorithm is useful for balancing the load over the S switch banks/AWGRs and minimizing the crosstalk in AWGRs used (passive approach). In the simulation conducted for the active approach, we used only one switch bank. Thus, we will describe the switch bank/AWGR assignment steps using the AWGR term used in the passive approach. However, the same steps can be used for the switch banks in the active approach if $S > 1$.

The steps of the round-robin assignment algorithm can be summarized in the following steps:

1. $AWGR_scan_pointer = last_assigned_switch \oplus 1$; where \oplus is modulo- S addition. When the pointer reaches S , it resets again. The value of the $AWGR_scan_pointer$ is the switch index where the scheduler starts to check for connection possibility. We will refer to this switch in next steps as “switch under test”
2. The scheduler searches using first-fit algorithm for free FSR (free FSR here means that it is not used by the source and destination pods in another connection) to make the connection through the “switch under test”. The FSR assignment according to first-fit algorithm will be shortly discussed in the next subsection. If a free FSR is successfully found, the scheduler sends a response (acknowledgment) to the source

pod with the switch index and the free FSR index allocated. Also, the scheduler will update the *last_assigned_switch* value with *AWGR_scan_pointer* value.

3. If the AWGR does not have a free FSR for this connection, the *AWGR_scan_pointer* is updated as follows: $AWGR_scan_pointer = AWGR_scan_pointer \oplus 1$.
4. If $AWGR_scan_pointer = last_assigned_switch$, then the scheduler fails to setup a connection and the request is buffered in a queue dedicated to this source and destination pair. During the time the request is buffered in the scheduler, the data burst constructed at the source pod stays buffered in the assembly buffer until contention resolution is achieved.
5. If $AWGR_scan_pointer \neq last_assigned_switch$, then the steps will be repeated from step 2.

2.5.1.2 Wavelength/FSR allocation

For wavelength/FSR allocation, the scheduler uses first-fit algorithm where the scheduler searches the wavelengths/FSRs in a fixed order to find a free wavelength/FSR at the AWGR to which the *AWGR_scan_pointer* points. We will use the term “channel” to represent the wavelength in the active approach and the FSR in the passive approach. The steps of the first-fit allocation algorithm can be represented in the following:

1. The scheduler starts with the first channel at the switch bank/AWGR, i.e. $channel_scan_pointer = 0$.
2. The scheduler checks if the source pod is using this channel in another connection. Also, the scheduler checks if the destination pod receiver (O/E) dedicated for this channel is busy in another connection. If this FSR is free on both sides, the scheduler can successfully establish the connection at this channel under test.
3. If this channel is not free, the *channel_scan_pointer* is updated, i.e. $channel_scan_pointer = channel_scan_pointer + 1$
4. If $channel_scan_pointer = k$, then the scheduler fails to make the connection through this switch bank/AWGR.
5. If $channel_scan_pointer \neq k$, then the steps will be repeated from step 2.

2.5.2 Contention resolution and OBF algorithm

In this approach, the contending requests are kept in queues within the scheduler according to the source and destination pods. The corresponding data bursts are kept in the electronic assembly buffers at the source pod until the source pod receives the response from the scheduler. If the assembly buffer is full, the oldest burst in the buffer is discarded. Each request (control packet) has a field that carries information about the size of the corresponding data burst. The scheduler uses this piece of information to keep track of the buffer size so that it discards the corresponding request when the buffer is full. Also, this field can be used by the scheduler to predict the reservation time for this burst transmission.

In the active approach, the OBF scheduling algorithm is used for these contending bursts upon connection release. The scheduler searches for the oldest request at the head of line (HOL) of the contending requests queue that can be served and send a response for the corresponding source pod for transmission.

In the passive approach, the OBF scheduling algorithm is slightly modified to speed up the operation and allow for two connections at maximum upon connection release rather than one connection in the active approach. The scheduler can have several processors equal to twice the number of pods. These processors share the status of the switches in the datacenter. When a connection is released, the processor related to the input port checks the head of the queues that holds requests from the corresponding source pod to different destinations. It checks if any of the requests can be served according to the status of the output links of the AWGR or, in other words, which O/E corresponding to the recently freed FSR and AWGR at the destination pods. The scheduler gives priority to the oldest request. This priority will be given to all requests in this queue and only one response will be sent to the source pod. The source pod will empty the corresponding assembly buffer and encapsulate all bursts into one jumbo burst. This processor at the scheduler will set this served queue (if any) with the least priority for this input port.

The procedure used by the input port processor is also done by the processor related to the output port. This time the processor checks the requests at the head of the queues that holds requests from different sources to this specific destination. Among those requests, it searches for the ones that can be served by the currently released O/E. It gives priority to the queue that holds the oldest request at HOL. The scheduler will send a response to the corresponding source pod that has sent this request. The source pod will empty the corresponding assembly buffer and

encapsulate all bursts in one jumbo burst. The processor will then set this queue (if any) with the least priority for this output port.

2.6 Simulation results and discussion

2.6.1 Active approach simulation assumptions, results and discussion

Throughout simulations done using OPNET Modeler 16.1, we assume that the traffic introduced by each of the 10 pods ($N = 10$) to the core layer, with one switch bank ($S = 1$) that includes two switch fabrics ($w = 2$), has the same parameters. In addition, we consider that Ethernet packets arrive to the pod interface, such that packet inter-arrival times follow a lognormal distribution [96]. The data rate per wavelength is assumed to be 10 Gb/s. Finally, we assume that the packet length is exponentially distributed with a mean packet length of 850 bytes [96]. However, in the Ethernet standard, the packet length is limited by minimum and maximum lengths of 68 bytes and 1526 bytes, respectively. When a generated packet has a length less than the minimum packet length or greater than the maximum packet length, its length is adjusted to follow the Ethernet standard. The burst length threshold (B_{th}) used in the burst assembly algorithm is 8500 bytes, and the time threshold (T_{th}) is 50 μ sec. The shape parameter is $\sigma = 0.8326$, which indicates low degree of self-similarity, where $H \approx 0.51$. The simulation considers three different cases; the network while not using a contention resolution technique, the network while using infinitely long buffers to achieve 100% throughput, and the network while using buffers with a size limit.

Fig. 2.5 shows the packet transmission success probability versus traffic in the three cases. It is obvious that the use of buffers even with a size of a few kilobytes enhances the network performance considerably with respect to the case of not using a contention resolution technique. Thus, introducing a small buffer in the transmitter reduces the number of needed wavelengths (transceivers) to achieve the required quality of service (QoS), which reduces the system complexity in addition to reduction in power consumption. This enhancement in the network performance is at the expense of introducing some latency to the burst transmission in addition to the burst assembly latency for burst construction. Fig. 2.6 shows the latency introduced due to the use of buffers for contention resolution along with the latency introduced by the burst assembly. One curve shows the assembly latency for the employed hybrid assembly algorithm with B_{th} and T_{th} indicated above. The other four curves show the latency due to additional

buffering for contention resolution with different buffer sizes. It is obvious that for large buffer size, the latency introduced by contention resolution is dominant at high traffic scenarios, while at low traffic scenarios, the latency due to burst assembly is more dominant. This is because at low traffic scenarios the time taken for burst construction is relatively high. It is also worth mentioning that for the 25 kB buffer limit the total latency due to burst assembly and due to contention resolution is still less than the time threshold used by the hybrid assembly algorithm.

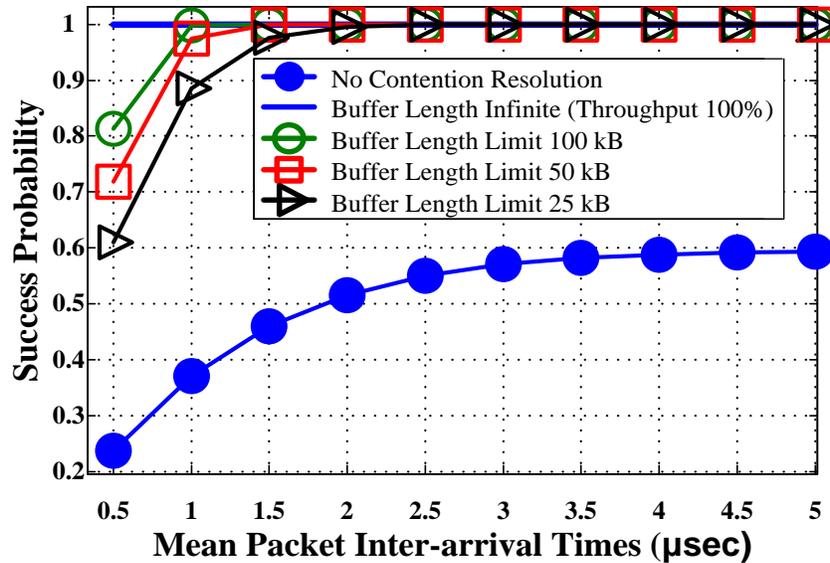


Fig. 2.5 Success probability versus mean packet inter-arrival times for different buffer limits (no contention resolution, infinite buffer limit and finite buffer limit).

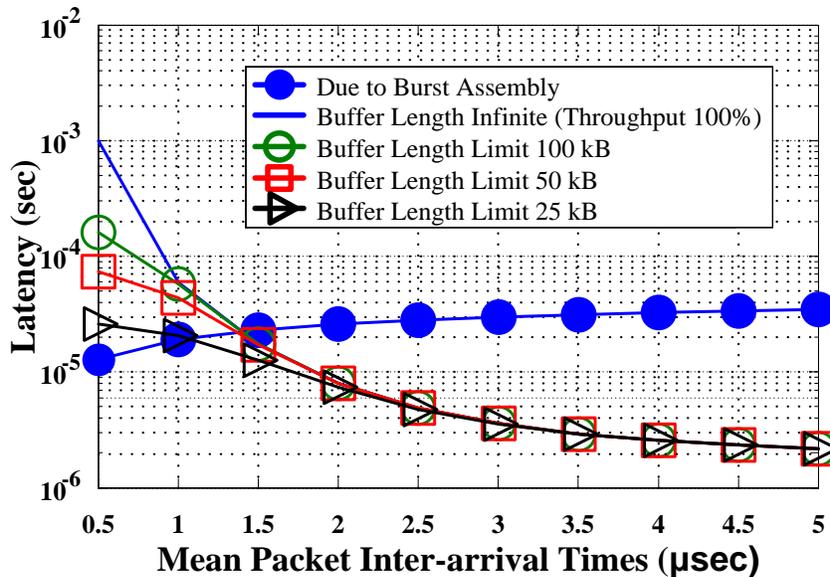


Fig. 2.6 Latency versus mean packet inter-arrival times (latency due to single burst construction is common for all buffer limits, latency due to contention resolution for different buffer limits).

Next, Fig. 2.7 shows the buffer utilization in the two cases; the 100% throughput case and the limited buffer size case. It is clear that at low and moderate traffic scenarios, the utilization of the low buffer sizes is more than larger buffer sizes. However, at high traffic scenarios, the utilization slope starts to change such that the 100 kB buffer has the highest utilization. This is attributed to the difference in latency order which is more apparent at high traffic scenarios making the same burst take a long time in the buffer to get served with respect to the case of the least buffer size.

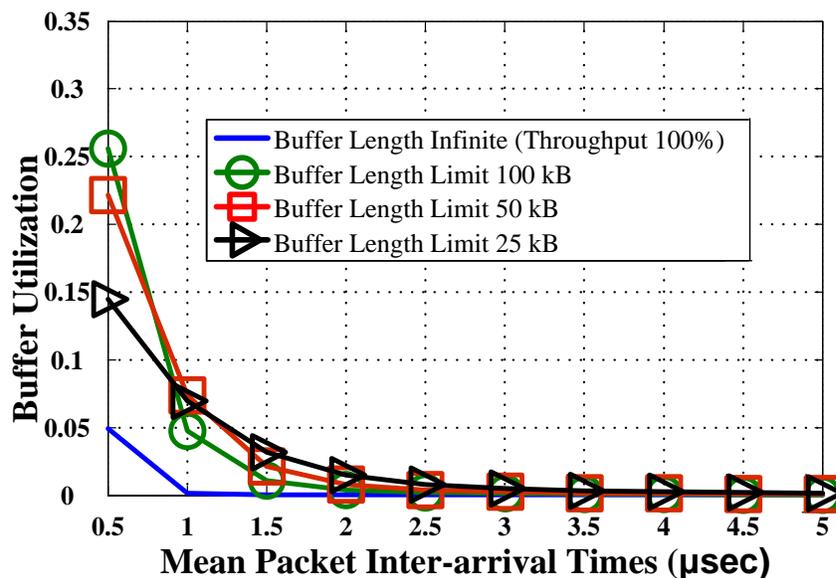


Fig. 2.7 Buffer utilization versus mean packet inter-arrival times for different buffer limits.

2.6.2 Passive approach simulation assumptions, results and discussion

In order to assess the effect of self-similarity of the incoming traffic, we compare the performance of the network in the case of packet switching for two traffic scenarios: self-similar with lognormal distributed inter-arrivals (at different values of the shape parameter σ), and non-self-similar Poisson traffic (inter-arrival times follow exponential distribution). The number of pods is 32 where each pod has 2 transceivers and the number of switches is 1. The assembly thresholds are $B_{th} = 180,000$ bits and $T_{th} = 50$ μ sec at $\sigma = 0.8326$ and $\sigma = 3$, but T_{th} at $\sigma = 3.7169$ is 70 μ sec. This change in time threshold at $\sigma = 3.7169$ gives better throughput performance compared to that when $T_{th} = 50$ μ sec. Buffer size limit of single electronic burst assembly buffer is 100 kB. Fig. 2.8 and Fig. 2.9 show clearly the effect of increasing the self-

similarity where the performance of the network degrades with this increase. Each value of the shape parameter (σ) has a corresponding value for the Hurst parameter (H). There are several methods for Hurst parameter calculation introduced in literature [100]. We used the absolute value method using a MATLAB function by passing the traces exported from OPNET as inputs to this function [101]. At $\sigma = 0.8326$, $H \approx 0.51$ which is the least degree of self-similarity, the performance is close to that of the exponential distribution. For burst switching, the performance of the network also degrades with increasing H . Burst switching at both low traffic scenarios and low degree of self-similarity is not effective because of latency induced from burst assembly. However, the burst switching is effective at high traffic scenarios where there are enough packets to affect the traffic characteristics as shown in Table 2-1. The self-similarity after burst assembly has been reduced especially at high traffic values.

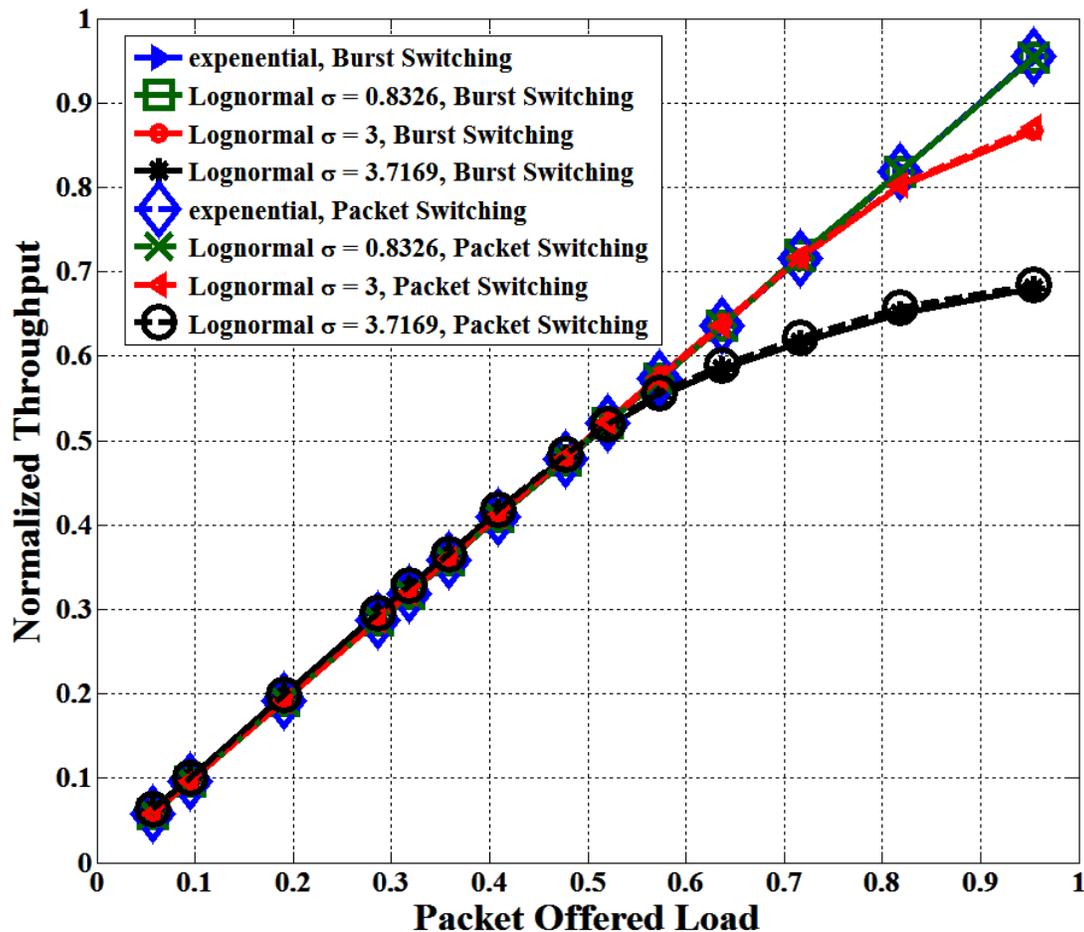


Fig. 2.8 Normalized throughput versus packet offered load for both burst and packet switching in both cases; exponential case and lognormal case (at different degrees of self-similarity).

Table 2-1 Hurst parameter of traffic at different scenarios

	Traffic Load	Packet Generator	After Burst Assembly	At Scheduler
$\sigma = 3$	High Traffic	≈ 0.75	0.662	0.582
	Medium Traffic	≈ 0.75	0.711	0.627
	Low Traffic	≈ 0.75	0.737	0.675
$\sigma = 3.7169$	High Traffic	≈ 0.86	0.786	0.621
	Medium Traffic	≈ 0.86	0.806	0.662
	Low Traffic	≈ 0.86	0.848	0.711

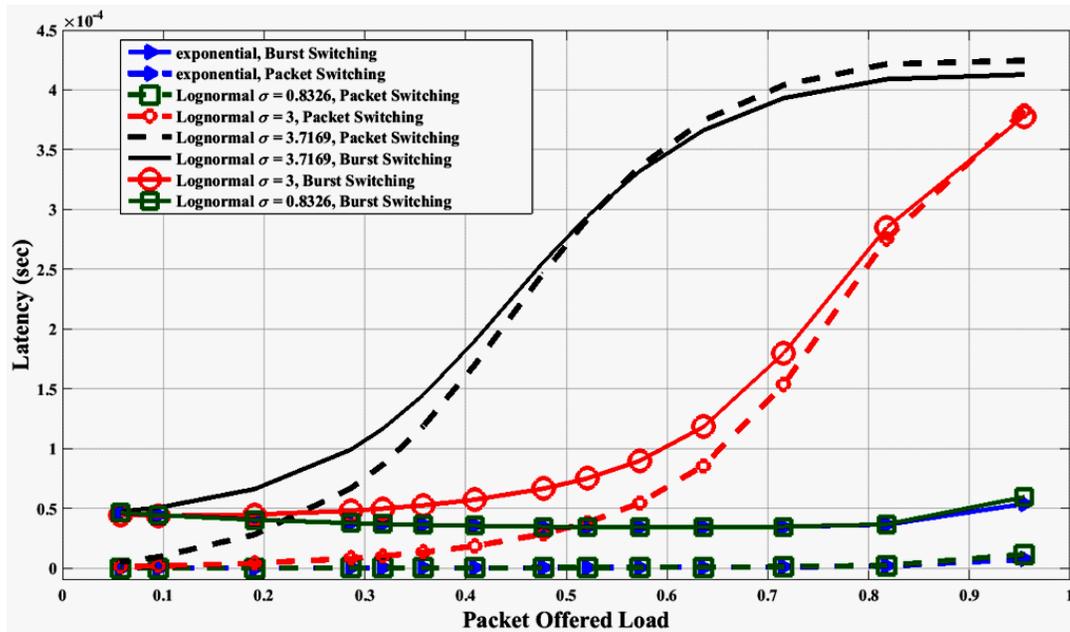


Fig. 2.9 Latency versus packet offered load for both burst and packet switching in both cases; exponential case and lognormal case (at different degrees of self-similarity).

From Fig. 2.9, it is obvious that the main advantage of burst switching appears in the latency especially when the latency due to contention resolution is dominant over the latency due to burst assembly. At high traffic and high degree of self-similarity, the latency due to contention in burst switching is less than that of packet switching. However, when we add the latency due to assembly (which is the price paid to reduce the traffic self-similarity) to the latency due to contention, the total latency approaches that of packet switching though burst switching still

slightly achieves less latency for high traffic loads. Thus, the main advantage of the OBS over OPS in this case is the reduction of the computational complexity at the scheduler due to the reduction in the connection requests.

In the next test scenario, we have the same datacenter under test where $N = 32$, $S = 1$, and $k = 2$, but we only study an example for the medium degree of self-similarity where $\sigma = 3$ ($H \approx 0.75$ at the source pod packet generation) while varying the buffer size limit of the electronic assembly buffers. Fig. 2.10 shows the network throughput versus packet offered load at different limits of assembly buffer size. When the assembly buffers are used in contention resolution, it can be seen as moving the VOQs from the core switch inputs to the transmitters. The use of buffers enhances the network performance at the expense of introducing some latency to the burst transmission in addition to the burst assembly latency for burst construction. Thus, increasing the assembly buffer size in the transmitter will reduce the number of needed FSRs (transceivers) to achieve the required quality of service (QoS) in terms of throughput, which reduces the system complexity in addition to reduction of power consumption.

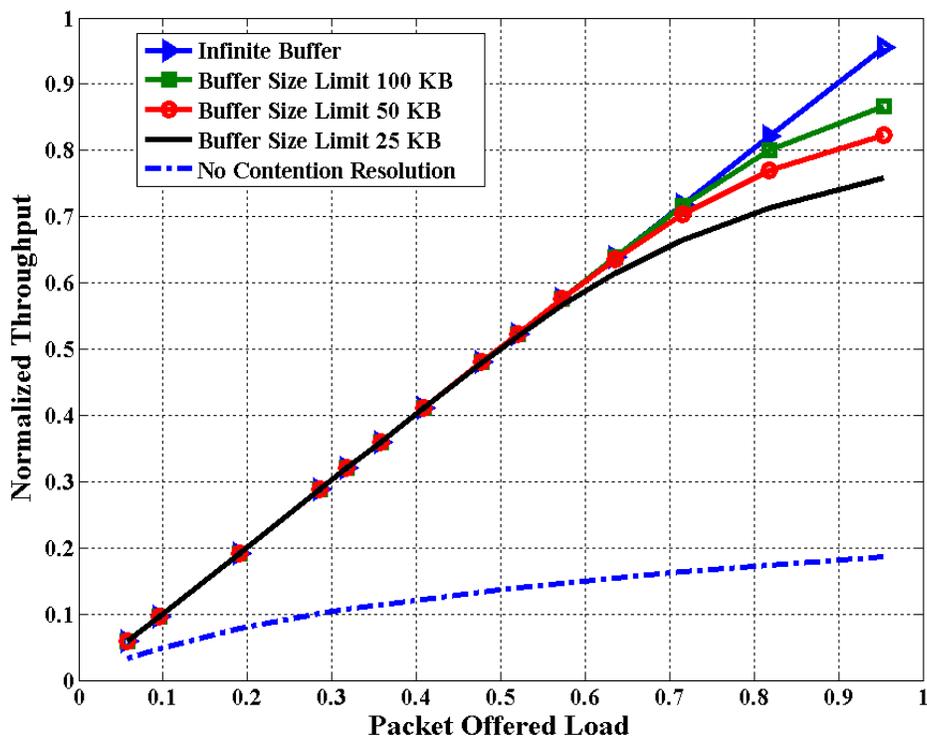


Fig. 2.10 Normalized throughput versus packet offered load at different buffer size limits.

Fig. 2.11 shows the latency introduced due to the use of buffers for contention resolution along with the latency introduced by the burst assembly (which can be seen in case of no contention resolution). It is obvious for large buffer size that the latency introduced by contention resolution is dominant at high traffic scenarios, while at low traffic scenarios the latency due to burst assembly is more dominant. This is because at low traffic scenarios the time taken for burst construction is relatively high.

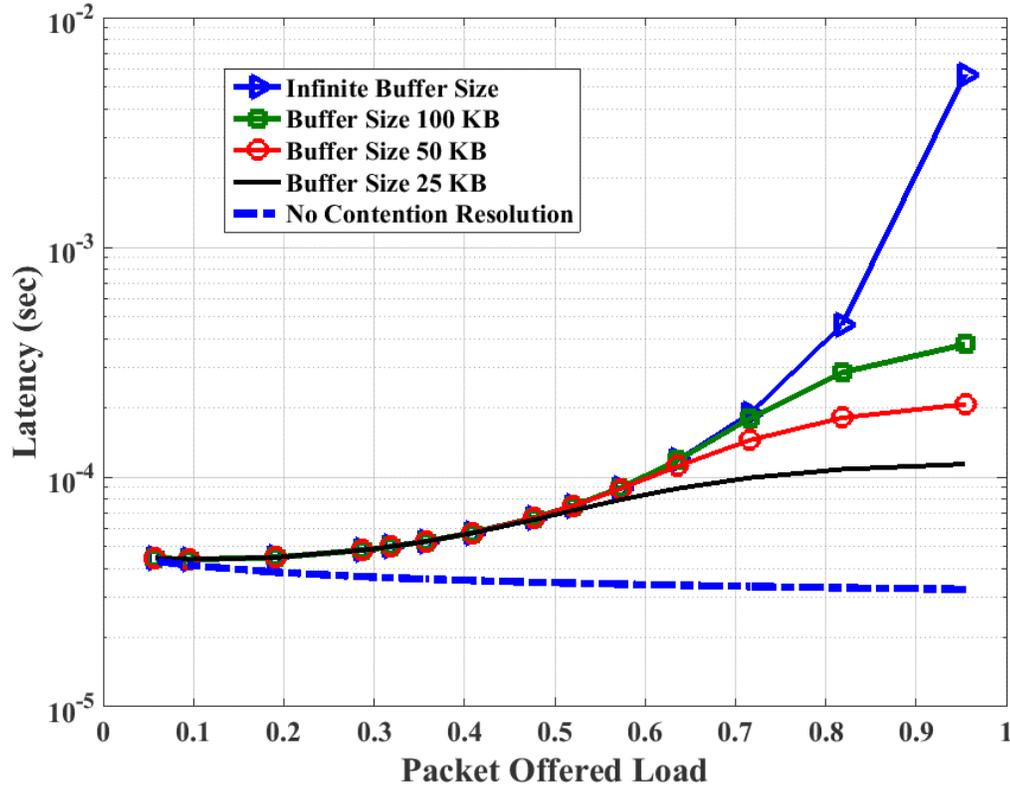


Fig. 2.11 Latency versus packet offered load at different buffer size limits.

Finally, we illustrate the tradeoff between network resources scale up and increasing the buffer size. In other words, we determine how much the network resources, i.e., number of AWGRs (S) and FSRs (k), need to be increased in case no contention resolution is used in order to achieve the same throughput as the case when contention resolution is used. Fig. 2.12 shows the throughput (the served portion of the traffic) versus the traffic load per pod, i.e., the load per link between a pod and an AWGR, in case of using the contention resolution strategy. In case of absence of the contention resolution strategy, the available bandwidth is scaled up by scaling up the network resources (AWGRs and FSRs). The normalization here differs to match the case of

using the contention resolution strategy for fair comparison. The required network resources, in case of absence of the contention resolution strategy, are eight times the required network resources, in case of using the contention resolution strategy with buffer size 50 kB, to reach almost the same throughput. However, when the network resources are just doubled, in case of using the contention resolution strategy with buffer size 50 kB, the throughput matches the case of using infinite buffer.

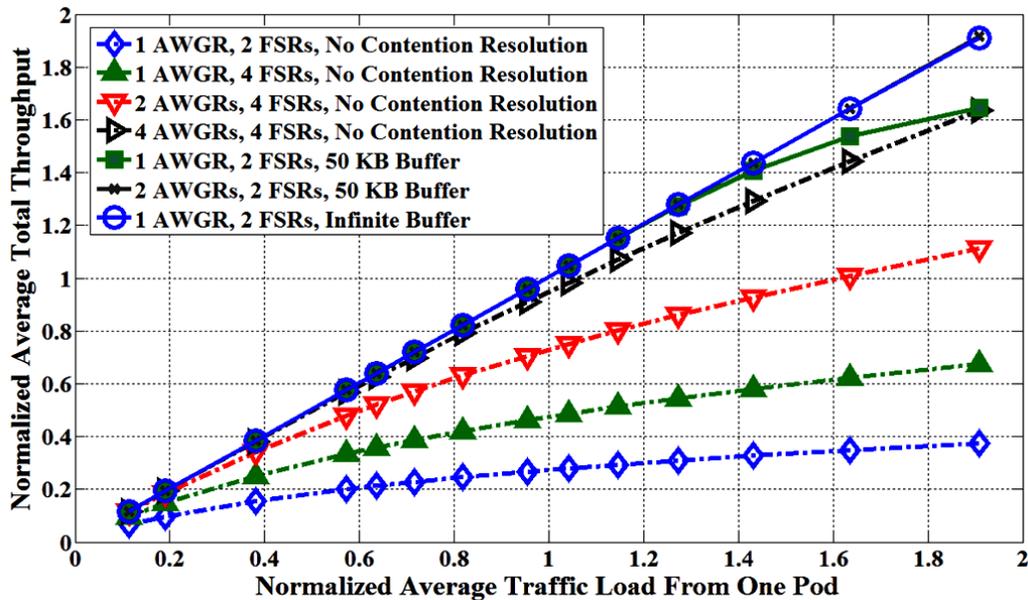


Fig. 2.12 Normalized average throughput (average over all switches going to a single pod and normalized w.r.t. data rate) versus normalized average traffic load offered from a single pod.

2.7 Conclusion

We proposed a new contention resolution strategy in OBS networks for intra-datacenter networking. The use of contention resolution enhances the network performance and gives the advantage of reducing the system complexity, and power consumption by reducing the required bandwidth, i.e., number of wavelengths. This leads to a reduction in the number of switches and transceivers and leads consequently to a reduction in power consumption. The tradeoff between latency, throughput, complexity, and power consumption can be made by adjusting the appropriate thresholds in the burst assembly, and choosing the appropriate buffer limit and system bandwidth.

In addition, we have proposed an AWGR-based switching mechanism that allows for parallel transmission from each pod at the same input port. The proposed architecture has the advantage

that the core switch is passive and the fixed continuous wave lasers are replaced by fast tunable lasers at the transmitter to be used for burst transmission and for wavelength routing at the same time. Furthermore, this mechanism avoids the buffering problem of optical switches by using the electronic burst assembly buffers at the transmitter instead of VOQs at the input of each switch for contention resolution.

Regarding the control plane, the centralized scheduler is realizable in case of intra-datacenter networking because of the locality of its entities. The centralized scheduling is used to control connections between pods not at the granularity of servers. In addition, the asynchronous service based on events and the separation between the scheduling for new requests and the scheduling of the contending requests, allows for parallel distributed processing at the scheduler while sharing the status of switches.

The advantage of using OBS is the reduction or smoothing of self-similar characteristics of the network traffic and the reduction of the number of requests to be handled by the scheduler. Also, the flexible size of the burst allows forming a jumbo burst that contains several bursts when the request of the head of line burst is served. This also reduces the scheduling time at high traffic.

The tradeoff between OBS and OPS changes according to the degree of self-similarity of the traffic. However, both can be simultaneously implemented as a hybrid solution, especially if software defined network (SDN) is used over the centralized scheduler, where traffic analysis can take place. Based on traffic analysis/prediction, the network can switch from OBS to OPS or vice versa by adjusting B_{th} in the assembly algorithm. This hybrid approach would benefit from both OBS and OPS depending on the traffic scenario.

Chapter 3

Self-Homodyne System for Datacenter Interconnects in Intra-Datacenter Networking

3.1 Overview

Cloud computing services and Internet of Things (IoT) impose a huge traffic growth in intra-datacenter networking. This in turn pushes the data rate requirements of DCIs in intra-datacenter networking to unprecedented extent. In the last few years, 100 Gb/s Ethernet has been standardized for short reach applications over 2 km and 10 km to use four wavelengths in the 1300 nm wavelength window with 800 GHz channel spacing. Each wavelength carries 25 Gb/s on-off-keying (OOK) signal in an IM/DD system [102]. For ~100 m fiber links, four fiber lanes are used to enable 25 Gb/s transmission in the 850 nm wavelength window with common form-factor pluggable transceivers (CFP and CFP2) [102].

In order to limit the number of used wavelengths and fiber lanes, several approaches have been proposed to increase the capacity per wavelength on a single polarization in an IM/DD system including 4-level pulse amplitude modulation (PAM-4) [37, 42-44, 103], discrete multi-tone technique (DMT) [47, 50, 51, 104], and multi-band carrierless amplitude phase (multi-CAP) modulation [47, 49, 105, 106]. Advanced solutions are also proposed where polarization-division multiplexed (PDM) intensity modulation transmission is achieved using the Stokes space representation of the optical signal to realize Stokes vector direct detection receiver [78, 83]. Also, a self-coherent approach for short reach applications has been proposed in [81] where a complex modulated signal using orthogonal frequency division multiplexing (OFDM) is polarization multiplexed with a tone originating from the same laser at the transmitter. The Stokes vector direct detection receiver is also adopted to retrieve the OFDM modulated signal. Most of these approaches are enabled by high speed DACs, ADCs and digital signal processing (DSP) at the transmitter and receiver sides.

Recently, efforts are directed towards standardizing of 400G networks for short reach applications (<40 km), i.e., IEEE 400GbE Task force [35]. The proposed solution for 500 m fiber

links is four fiber lanes, where each lane operates at 50 Gbaud using PAM-4 modulation format. On the other hand, the adopted solution is a WDM transmission using 8 wavelengths for 2 km and 10 km on 800 GHz grid where each wavelength carries ~25 Gbaud signal using PAM-4 modulation format. In addition, it is expected that standardization will be required for 800 Gb/s and 1.6 Tb/s transmission for short reach applications [36]. This motivates finding robust solutions that cope with the endless increase of data rate requirements of DCIs in intra-datacenter networking.

In parallel to capacity considerations while maintaining reduced system complexity, developing cost effective small form factor pluggable transceivers is desired for the massive number of DCIs in intra-datacenter networking. Thus, technology of integrated photonics is investigated to realize cheap small form factor solutions. As mentioned in 1.2.2, Si and InP platforms are useful hosts for integrated photonics devices, e.g., modulators and photodetectors [3, 4]. 112 Gb/s experimental demonstration using SiP modulator for short reach applications has been reported for PAM-4 transmission on a single polarization [39]. Also, PDM transmission in a Stokes vector direct detection system has been reported using SiP modulator and polarization multiplexed emulation [83]. In [107], SiP Stokes vector transmitter using IQ modulator polarization-multiplexed with a carrier is tested against SiP Stokes vector receiver on another chip. For InP modulators, experimental demonstrations have been reported for inter-datacenter networking using IQ modulation [52, 53, 58-60, 62]. The integration of the InP modulators with RF drivers and lasers has been demonstrated in [53] and [52], respectively. The InP modulators have the advantage of operating at low values of V_π with high 3-dB bandwidth relative to SiP solutions [63, 64].

In this chapter, we propose a self-homodyne system for next generation DCIs in intra-datacenter networking [108]. This enables the full utilization of the four degrees of freedom available for modulation in optical transmission. The main concept is to utilize the full-duplex fibers deployed in intra-datacenter networking to connect between optical transceivers to send the tone and the modulated signal from the transmitter to the receiver for coherent reception. Optical circulators are used to launch/extract the tone and the modulated signal to/from the full-duplex fiber. An InP DP-IQM, with 35 GHz 3-dB bandwidth and $2.5 V_\pi$, is used at the transmitter side for coherent transmission. This self-homodyne system has reduced Rx-DSP complexity compared to conventional coherent system approaching the Stokes vector direct

detection systems complexity. We report for the first time to the best of our knowledge 530 Gb/s transmission over 500 m fiber on a single wavelength using 53 Gbaud PDM-32QAM signal below the KP4 FEC threshold (bit error rate (BER) of 2.2×10^{-4}) [34]. In addition, we demonstrate a 448 Gb/s transmission over 2 km fiber using 56 Gbaud PDM-16QAM at BER of 8.5×10^{-6} which is below the KP4 threshold with more than an order of magnitude. Finally, transmissions over 10 km at bit rates of 320 Gb/s and 360 Gb/s are realized at symbol rates of 40 Gbaud and 45 Gbaud below BERs of 2.2×10^{-4} and 2.8×10^{-4} , respectively, using 16QAM modulation format.

The rest of this chapter is organized as follows: In section 3.2, we describe the principle of the proposed self-homodyne system. Section 3.3 is dedicated to the experimental setup description of the system emulation. The reduced DSP stack from the conventional coherent system is discussed in section 3.4. Finally, the experimental results are presented in section 3.5 and we conclude in section 3.6.

3.2 Principle of the proposed self-homodyne system

The main concept of any self-homodyne system is to send the complex modulated signal and a copy of the tone originating from the same laser multiplexed in a specific domain from transmitter to (Tx) to the receiver (Rx) to be used for coherent reception. For example, the tone can be sent on a different fiber core or mode in a space-division multiplexed system [109, 110]. Another approach is to send the tone polarization-multiplexed with the modulated signal as in Stokes vector direct detection (SV-DD) system [81]. This tone essentially replaces the local oscillator (LO) in a conventional coherent system. The benefit of this homodyne system is that the DSP for both frequency offset removal and phase noise mitigation can be omitted which reduces the DSP complexity for such system. Hence, power consumption will be reduced compared to conventional coherent systems, which is crucial for intra-datacenter applications.

The proposed self-coherent system uses the full-duplex fiber deployed in short reach applications by transmitting the modulated signal on one fiber and the tone on the other fiber. Fig. 3.1 shows a schematic of the proposed system including two transceivers. Each transceiver has a coherent transmitter and receiver (CRx) with the required digital-to-analog and analog-to-digital converters (DAC and ADC) to enable transmitter and receiver digital signal processing (Tx-DSP and Rx-DSP), respectively. We add two circulators at each transceiver side to use the

two fibers (colored in red and green in Fig. 3.1) of the full-duplex fiber as bidirectional fibers instead of unidirectional fibers in conventional systems. The circulators C1 and C3 couple the modulated signal from each transceiver to the full-duplex fiber for transmission, and extract the tone sent on the same fiber by the other transceiver for coherent reception. Likewise, circulators C2 and C4 are used to couple the tone from each transceiver to the full-duplex fiber, and extract the transmitted signal from the other transceiver. The tone path includes a polarization stabilizer to ensure having tone-signal beating on both state of polarizations. This polarization stabilizer can be realized using integrated optoelectronics using the polarization controller design proposed in [111, 112]. Matched fibers are used to match the delay between the signal and the tone to ensure that they have the same laser phase noise at CRx to be canceled during the tone-signal beating. In addition, matching the tone and the signal paths enables the system implementation using uncooled lasers where the drift (Δf) in the center frequency (f_c) of the laser affects the center frequency of the tone and the modulated signal simultaneously. This drift in the center frequency shall be omitted during the tone-signal beating process where the center frequency of the tone and the signal are the same after the drift ($f_c + \Delta f$).

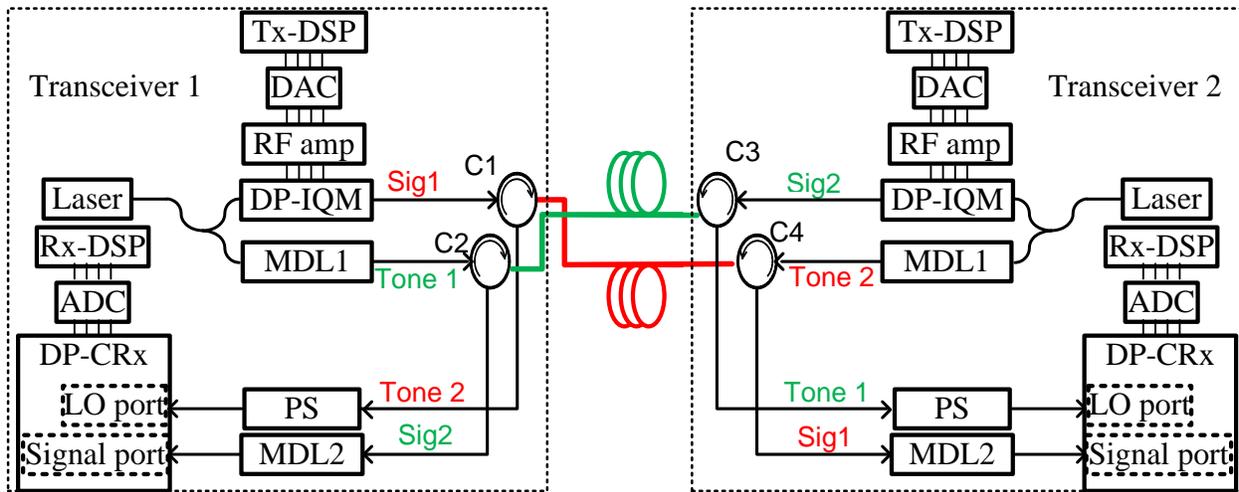


Fig. 3.1 Schematic of the proposed self-homodyne system (PS: polarization stabilizer, MDL: matched delay line).

3.3 Experimental setup

Fig. 3.2 shows the experimental setup used to emulate the system described in the previous section. To do this emulation, we connected the red fiber link shown in Fig. 3.1 between optical circulator C1 and optical circulator C2. This provides the effect of the counter propagating tone in the opposite direction of the optically modulated signal propagation direction.

The transmitter side, shown in Fig. 3.2, starts with an AC-coupled 8-bit DAC, operated at 84 GSps. The four differential outputs of the DAC are connected to a quad linear amplifier with 30 GHz 3-dB bandwidth. The quad linear driver generates four single-ended RF signals used to drive the fully packaged 35 GHz 3-dB bandwidth InP-based DP-IQM [59]. A 100 kHz linewidth external cavity laser (ECL) operating at 1565 nm with 15.5 dBm optical power is used as the origin optical source for the tone and the optical data signal. The continuous wave (CW) light from the ECL is split using 70/30 polarization maintained (PM) coupler to provide the DP-IQM with the input CW light through the 70% port.

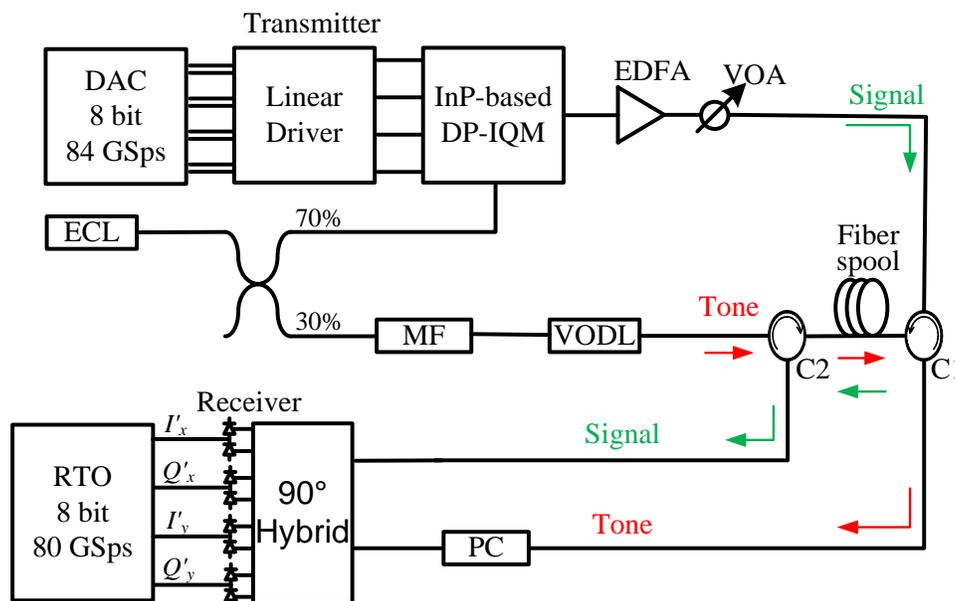


Fig. 3.2 Experimental setup for the proposed system emulation (MF: matching fiber, and PC: polarization controller).

After optical modulation, an Erbium doped fiber amplifier (EDFA) is used to boost the signal power level followed by a variable optical attenuator (VOA) which is mainly used for power adjustments to keep the received power at different symbol rates constant. The received signal power levels in the back-to-back (B2B) case, and at 500m, 2 km, and 10 km are 6.5 dBm, 6.3

dBm, 6 dBm, and 4.5 dBm, respectively, at the input ports of the optical hybrid. The amplified signal is then launched to the SMF 28e+ fiber spool through circulator C1. For the tone, the output of the 30% port of the PM coupler is connected to circulator C2 after matching the delay between paths to launch the tone to the fiber spool. Thus, the tone and the modulated optical signal are counter-propagating on the same fiber link. The fiber spools' lengths used in our demonstration are 500 m, 2 km and 10 km which are the standard lengths used in short reach applications with full-duplex fibers. In the B2B case, we replace the fiber spool with 30 cm patch cord, to keep the effect of the circulators and also maintain the matched paths that include the circulators' pigtail lengths.

At the receiver side, the tone and the modulated optical signal are extracted after propagation using circulators C1 and C2, respectively, for coherent reception. The tone and the modulated are then directed to the LO and signal input ports of the optical hybrid, respectively. The path of the tone includes a polarization controller, which replaces the polarization stabilizer in Fig. 3.1, to manually adjust the state of polarization of the tone. The tone power levels, after the PC, equal to 7.4 dBm, 7.25 dBm, 7 dBm, and 5.4 dBm in the B2B case and at 500 m, 2 km, and 10 km, respectively. The eight outputs of the optical hybrid are connected to four balanced photodetectors (BPDs) feeding 8-bit real time oscilloscope (RTO) with four 33 GHz input channels operating at 80 GSps for offline signal processing.

It should be noted that the EDFA is used to compensate for the 10 dB loss of the optical hybrid at the receiver and the 2 dB loss of the circulators set providing sufficient received optical power to the balanced photo-detectors where no transimpedance amplifiers follow. Hence, in a real system, the EDFA is not needed when high bandwidth transimpedance amplifiers (TIAs) are used after the balanced photodetectors following the optical hybrid. We used the EDFA at the transmitter side to avoid the induced optical noise when the input signal level is reduced by the circulators set loss besides the propagation attenuation loss, if the EDFA is deployed at the receiver.

3.4 DSP stack

Fig. 3.3 illustrates the offline DSP stack used at the transmitter and the receiver. The transmitter DSP (Tx-DSP) starts with N-QAM symbol generation followed by pulse shaping at two samples per symbol as shown in Fig. 3.3(a). The pulse shaping filter is a root raised cosine (RRC) pulse

shaping filter with roll-off factor dependent on the operating symbol rate. The roll-factor values used at symbol rates of 28 Gbaud, 35 Gbaud, 40 Gbaud, 45 Gbaud, 50 Gbaud, 53 Gbaud, and 56 Gbaud are 1, 0.82, 0.6, 0.42, 0.28, 0.2, and 0.15, respectively. Next, the pulse shaped signal is resampled at DAC sampling rate of 84 GSps. After resampling, clipping and non-linear compensation are performed where the transfer function of the used modulator [59] is inverted to insure equally spaced QAM constellation points after optical modulation. Then, we pre-distort the signal using an experimentally optimized digital pre-emphasis filter to partially equalize the limited bandwidth of the transmitter components [59].

The receiver DSP (Rx-DSP) in the proposed self-homodyne system is a reduced version of the conventional coherent Rx-DSP where frequency offset removal and phase noise mitigation are omitted. The chromatic dispersion (CD) compensation is also omitted since short fiber links are deployed and the compensation can be included within subsequent adaptive filtering. As shown in Fig. 3.3(b), Rx-DSP starts with IQ imbalance compensation [75] followed by resampling where the rest of the Rx-DSP operates at double the symbol rate. It should be noted that we use the same reference clock for the DAC and the RTO due to the co-locality of the system. Thus, the clock recovery DSP block is not required in our demonstration. Next, synchronization is done to find the training symbols within the captured frame for initial polarization tracking. The polarization tracking is performed using 4×4 real valued multiple-input-multiple-output (MIMO) DSP block [60] shown in Fig. 3.3(b). Initially, we use training symbol least mean squares algorithm (TS-LMS) for initial polarization tracking. Then, we switch to decision directed least mean squares algorithm (DD-LMS) after convergence for steady state operation. In addition to the polarization tracking, the 4×4 real valued MIMO includes adaptive filter taps that (1) remove any residual inter-symbol interference (ISI), (2) include the RRC matched filtering, and (3) compensate for CD induced from short distance propagation. It is worth noting that if the operating wavelength is in the O-band window, the CD effect will be minimal, which is the case in most of the DCIs used for intra-datacenter networking.

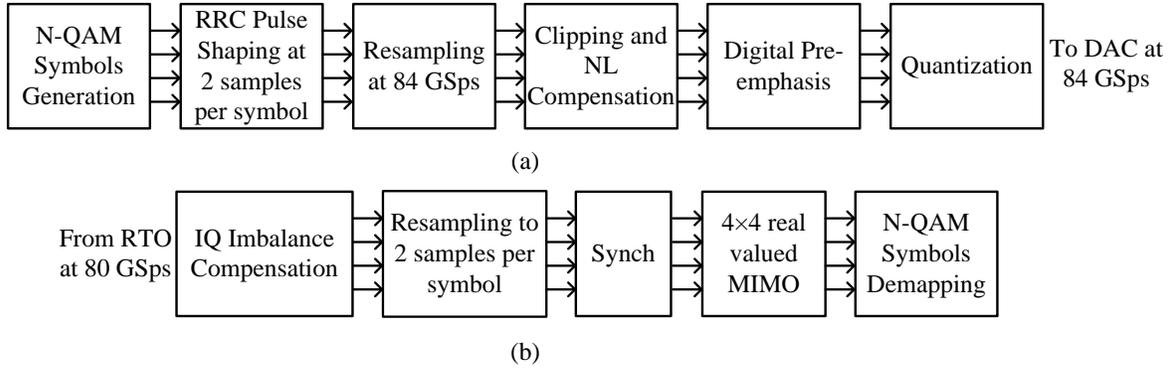


Fig. 3.3 DSP stack for a self-homodyne system; (a) Tx-DSP, and (b) Rx-DSP.

3.5 Results and discussion

Fig. 3.4 shows the BER versus distance at different symbol rates using 16QAM modulation format. Below 10 km which is the longest distance that uses full-duplex fibers according to the IEEE 400GbE task force [35], we are able to achieve data rate of 448 Gb/s at a BER 8.5×10^{-6} below the KP4 FEC threshold using a symbol rate of 56 Gbaud on a single wavelength in the B2B case. Also, if we consider the hard decision forward error-correcting threshold (HD-FEC) of 3.8×10^{-3} at 10 km, 400 Gb/s payload rate after overhead removal can be achieved. For 40 Gbaud, 45 Gbaud, 50 Gbaud, and 56 Gbaud 16QAM signals, data rates of 320 Gb/s, 360 Gb/s, 400 Gb/s, and 448 Gb/s are achieved below BER of 2.2×10^{-4} , 2.8×10^{-4} , 4.1×10^{-4} and 7.5×10^{-4} at 10 km, respectively.

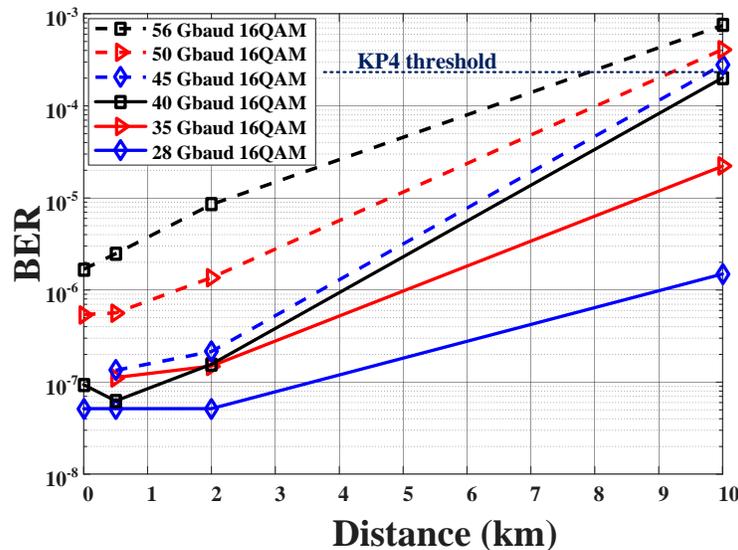


Fig. 3.4 BER versus distance at different symbol rates using 16QAM modulation format.

The main reason for the system performance degradation at 10 km is the stimulated Brillouin scattering (SBS), which is an inelastic scattering mechanism that occurs when a continuous wave (CW) light propagates in the fiber with high power [7]. The backward reflection (in direction from C1 to C2) due to SBS generated by the tone propagation from C2 to C1 pollutes the signal propagating from C1 to C2 at optical frequency $f_c - 10.73$ GHz where f_c is the optical frequency of the CW tone in GHz. This backward reflection due to SBS is then downconverted by coherent reception to ~ 10.73 GHz polluting the spectrum of the received signal due to the unwanted tone at 10.73 GHz which has ~ 55 MHz bandwidth as shown in Fig. 3.5. This is another important reason why we used the EDFA at the transmitter not at the receiver to avoid amplifying the SBS reflection effect. In a real system, this can be totally avoided in the real system when transceiver 1 and transceiver 2, shown in Fig. 3.1, use different wavelengths (separated by more than the Brillouin shift of 11 GHz) such that the counter-propagating tone from transceiver 2/1 generates a reflected light away from the data signal from transceiver 1/2. Since our purpose is to demonstrate the idea of implementing the self-homodyne systems in short reach DCIs, we believe that the current emulation, though it is SBS limited, adequately proves the viability of the system.

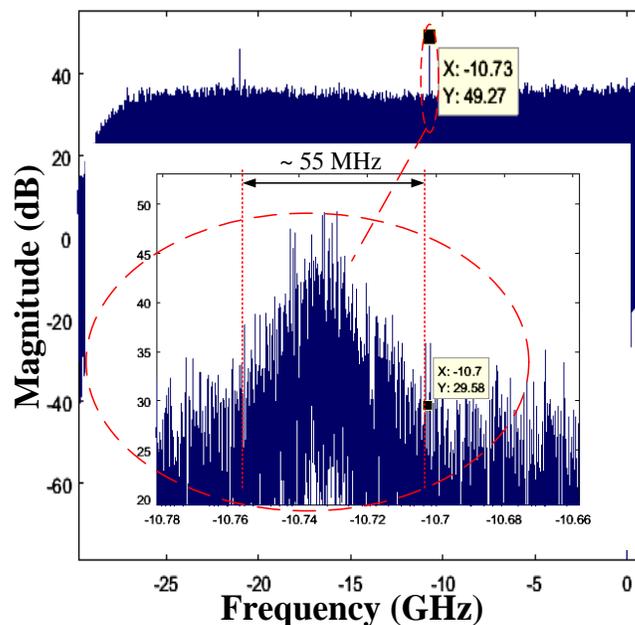


Fig. 3.5 56 Gbaud 16QAM spectrum from the captured data at the receiver.

Fig. 3.6 shows the BER versus distance at different symbol rates using 32QAM modulation format. At 500 m, BERs of 1.78×10^{-4} and 3.2×10^{-4} are achieved with symbol rates of 53 Gbaud and 56 Gbaud at 32QAM modulation format, respectively. This shows how the proposed system enables for very high data rate that exceeds 0.5 Tb/s achieving 530 Gb/s on a single wavelength at 500 m. At 2 km, 500 Gb/s and 450 Gb/s are achieved at BERs of $\sim 2.6 \times 10^{-4}$ and $\sim 1 \times 10^{-4}$ using 50 Gbaud and 45 Gbaud signals with 32QAM modulation format, respectively.

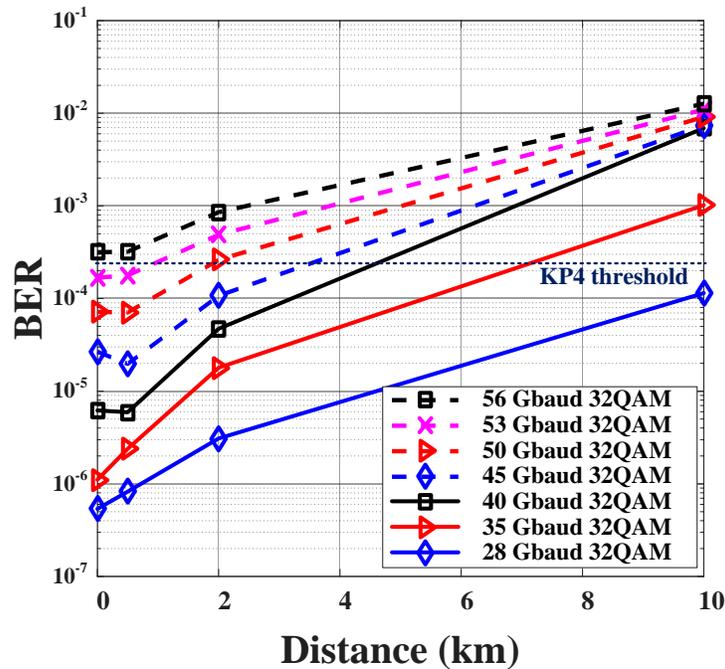


Fig. 3.6 BER versus distance at different symbol rates using 32QAM modulation format.

It is expected that the paths of the signal and the tone are not exactly matched at the receiver. Thus, in Fig. 3.7, we investigate the effect of this mismatch on the system performance for a 56 Gbaud 16QAM signal at 2 km. BER versus the mismatch length curves are generated by adding patch cords at different lengths in the tone path. It can be seen that the system performance is not affected by a mismatch less than 2 m. Phase noise mitigation is required to compensate for a mismatch more than 2 m as in [113] without the second order phase locked loop (PLL). It should be noted that the performance degradation due to path mismatch depends on the linewidth of the used laser, where the use of a laser with higher linewidth makes the system performance more sensitive to the mismatch between the tone and the signal paths. Also, another problem for the

path mismatch is the frequency drift of the laser, especially if the target is to realize uncooled systems. However, results show that the system performance is immune to the path mismatch to a reasonable extent given the current laser technology.

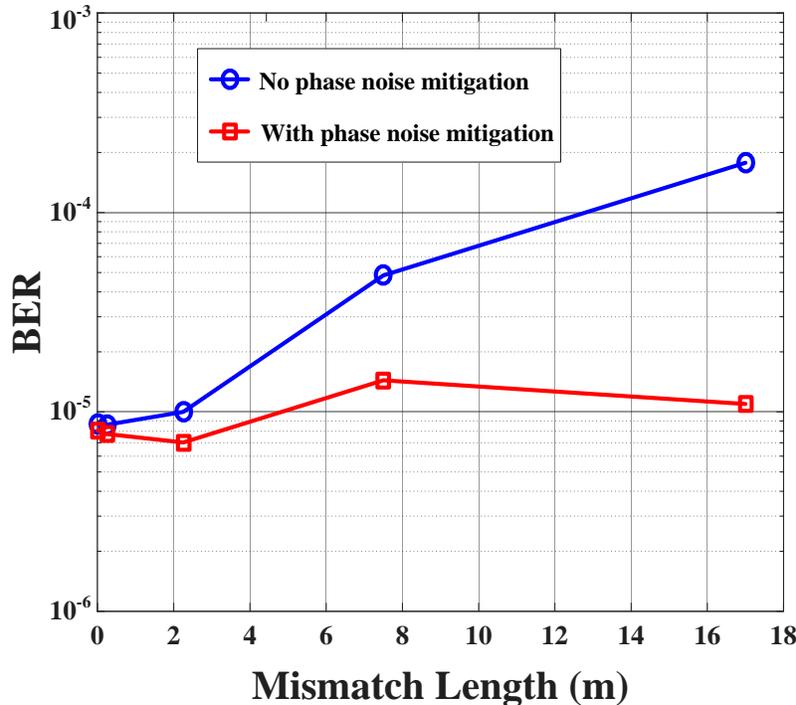


Fig. 3.7 BER versus path mismatch length between the tone and the modulated signal using phase noise mitigation and without using phase noise mitigation.

Finally, we show the effectiveness of having digital pre-emphasis at the transmitter in case of 56 Gbaud 16QAM signal at 2 km in Fig. 3.8. The digital pre-emphasis filter is optimized at 121 taps. In order to reduce the number of taps, we apply Kaiser Window to smooth the energy of the filter taps before truncating them. It can be seen that the increase in the number of the digital pre-emphasis filter taps enhances the system performance up to approximately 17 taps where no improvement in the BER is achieved with further increases in the pre-emphasis taps. Also, it is worth noting that we can pick a suboptimal point where we reduce the number of filter taps at the transmitter and the receiver achieving the required performance. For example, if the required BER is 2.2×10^{-4} at 2 km, we can use several (Tx taps at 84 GSps, Rx taps at 112 GSps in case of 56 Gbaud) pair to achieve this performance such as (25, 23), (17,33) and (9, 45).

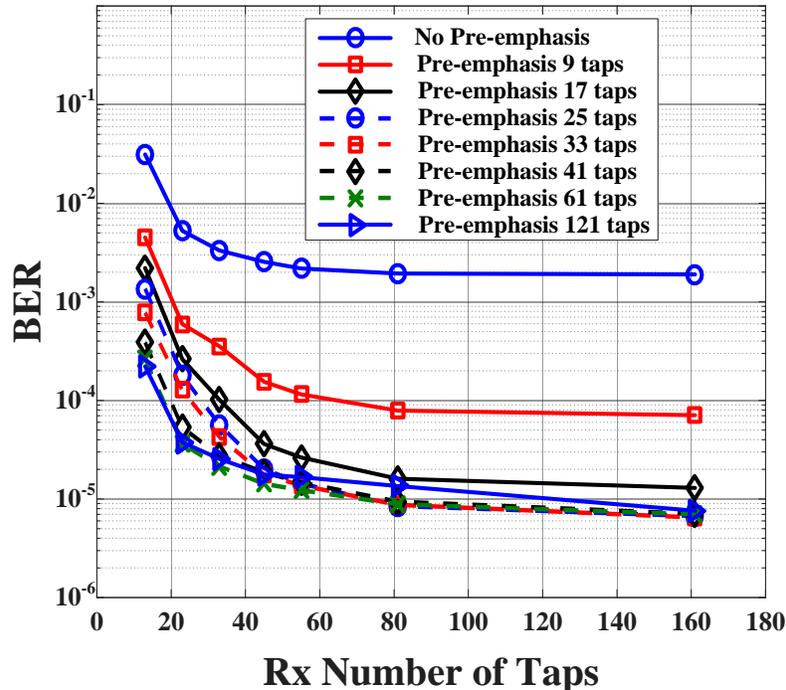


Fig. 3.8 BER versus receiver number of taps used in the real valued 4×4 MIMO at different number of taps used in the digital pre-emphasis at the transmitter.

3.6 Conclusion

We proposed a self-homodyne system for next generation DCIs in intra-datacenter networking and we experimentally demonstrated an emulation for the proposed system using integrated InP DP-IQM. The proposed system has several advantages beside the direct reduction of the Rx-DSP complexity compared to conventional coherent system by omitting both the frequency offset removal block and the phase noise mitigation process. In addition, the system has the advantage of achieving high bit rates with low bit error rates on a single wavelength compared to direct detection systems. This reduces the number of used components, e.g., lasers, modulators, and drivers, required by the direct detection system to achieve the same data rate. For example, the accepted solution in the 400GbE task force for 2 km reach is the use of 8 wavelengths; each carrying PAM-4 modulated signal operating at a symbol rate of ~25 Gbaud (8 lanes with 8 lasers, 8 modulators, 8 drivers, 8 PDs, and associated 8 TIAs). On the other hand, the proposed system achieves 448 Gb/s at 2 km using a single wavelength operating at symbol rate of 56 Gbaud with 16QAM modulation format (1 laser, 4 drivers, 4 BPDs (8 PDs), and associated 4 TIAs). In addition, it is a good candidate to realize a solution for uncooled lasers where the

proposed system achieves the required data rates on a single wavelength and is more immune to drifts in the laser frequency. These drifts in the center frequency are omitted during the tone-signal beating on the photodetectors of the BPDs since it affects the center frequency of the tone and the signal simultaneously provided that the tone and the signal paths are matched. This is a significant advantage compared to WDM solutions, where a robust design for a flat top Mux/Demux must be used to be able to tolerate the frequency drift by the uncooled lasers. Furthermore, the transmission in the C-band is possible since the compensation of the CD can be done within the adaptive filtering without a noticeable impact on the used number of taps, because the signal propagates over short distances in the SMF fiber. Finally, we reported, to the best of our knowledge, record breaking transmissions of 530 Gb/s, 448 Gb/s, and 320 Gb/s signals on a single wavelength over 500 m, 2 km, and 10 km of SMF fibers, respectively, below KP4 threshold using integrated InP modulator.

Chapter 4

High Speed Coherent Optical System Using InP Integrated Modulator

4.1 Overview

Evolution of internet services such as online video streaming and cloud computing has tremendously increased network capacity demand, especially for datacenter interconnects (DCIs) particularly for intra- and inter-datacenter applications. Optical coherent systems are used in inter-datacenter applications to achieve the highest possible capacity for reaches less than 1000 km. 400 Gb/s coherent transceivers based on high symbol rates and spectrally efficient modulation formats are anticipated to be widely deployed soon for short distance DCI applications to satisfy their requirements for high capacity and low cost [114]. One approach to achieve 400 Gb/s solution is to use high symbol rates and low order modulation formats. For example, a super-Nyquist filtered 110 Gbaud PDM-QPSK signal is used to reduce the signal bandwidth to 100 GHz [115]. Another approach is to use higher order modulation formats including 16QAM [116, 117] or 64QAM [118], and operate at lower symbol rates. All of these approaches require digital signal processing (DSP) at the transmitter and the receiver (some is done in the optical domain as in [117]). Another approach is to use either wavelength-division multiplexing, for example dual-carrier 32 Gbaud PM-16QAM [119], or time-division multiplexing [120].

Recently, several demonstrations have been reported for high-speed single carrier transmissions beyond 400 Gb/s solution [41, 121-124]. For example, WDM transmission of 72 Gbaud PDM-64QAM single carrier signals over 400 km of ultra large area fiber (ULAF) with hybrid Erbium doped fiber amplifier (EDFA)/Raman amplification at a bit error ratio (BER) of 3×10^{-2} was demonstrated in [121]. An optical filter was used to shape the spectrum to compensate for the limited bandwidth of the transmitter and a calibrated pre-distortion was implemented to mitigate for inter-symbol interference (ISI). Another approach is to operate at very high symbol rates by using digital bandwidth interleaving method to generate 140 Gbaud BPSK signal which can be translated to 140 Gbaud signal in coherent transmission [41].

Electrical time-division multiplexing (ETDM) is another approach to perform high speed transmission by increasing the symbol rate. For example, 120 Gbaud ETDM PDM-16QAM signal is demonstrated over 1200 km of TeraWave fiber spans with Raman amplification [122]. The real challenge in single-carrier high speed transmission especially using high modulation order formats is the overall limited bandwidth at the transmitter due to cascaded limited bandwidth electro-optic components [123]. In [123], authors have demonstrated 400G per channel WDM transmission with PDM-64QAM signal over 328 km of large area pure silica core fiber with hybrid EDFA/Raman amplification at BER of 5×10^{-2} . The data is generated using 88 GSps digital-to-analog converter (DAC) and 80 GSps analog-to-digital converter (ADC).

In parallel, the deployment of integrated photonics in several applications to achieve cost-effective, pluggable coherent transceivers has captured a lot of interest for manufacturers, researchers and designers [4]. The InP-based IQ modulator is a good candidate to realize integrated pluggable coherent transceivers with relatively higher 3-dB bandwidth at lower values of V_π compared to other integrated solutions [6, 52, 53, 61-64, 125]. Also, InP-based devices can be integrated with lasers [52], RF amplifiers [53] and SOAs [6]. Several single carrier transmission demonstrations have been reported that utilize InP-based IQ modulators operating over varying distances, data rates, and modulation formats. For example, the authors demonstrated long haul transmission at 32 Gbaud PDM-QPSK and PDM-16QAM with dual polarization emulation over 8000 km, and 960 km, respectively [52]. Also, a 28 Gbaud 64QAM modulated signal is transmitted over 40 km at a BER of 1.0×10^{-2} [62]. In [61], InP segmented Mach-Zehnder modulators (segmented MZMs) are used to realize a single polarization IQ modulator with DAC-free operation. Dual polarization emulation is done to demonstrate 32 Gbaud PDM-64QAM transmission over 80 km at BER of 9.1×10^{-3} .

In this chapter, we first demonstrate 400G transmission using a $2.5V V_\pi$ InP packaged dual-polarization IQ modulator (DP-IQM) with 35 GHz 3-dB bandwidth on a single carrier using 16QAM at 56 Gbaud, and 64QAM at 38 Gbaud, and 43 Gbaud [59]. These modulation schemes enable 400G transmission after overhead removal. To achieve this result, we implement DSP at the transmitter and the receiver, as will be discussed later, including transmitter frequency response equalization, non-linear compensation for MZM response, and other DSP algorithms. Also, an experimental comparison is performed between an InP-based and a LiNbO₃ based DP-IQM. This demonstration was enabled using a DAC operating at 65.7 GSps and 33 GHz real

time oscilloscope (RTO) operating at 80 GSps. Then, we further approach the bandwidth limit of the same InP DP-IQM and demonstrate payload rate flexible upto 600 Gb/s transmissions as a cost-effective solution for next generation DCIs and backbone networks. In particular, we demonstrate various symbol rates and modulation formats including 77 Gbaud PDM-32QAM (770 Gb/s), 84 Gbaud PDM-16QAM (672 Gb/s), 84 Gbaud PDM-8QAM (504 Gb/s), and 84 Gbaud PDM-QPSK (336 Gb/s), all enabled by a four-channel DAC operating at 84 GSps that drives the InP DP-IQM in the transmitter and two RTOs operating at 160 GSps in the receiver [60]. The transmission distance is 320 km, 960 km, 1980 km, and 4480 km for the four signals at a BER of 2×10^{-2} , respectively. The payload rate is about 600 Gb/s, 525 Gb/s, 400 Gb/s and 260 Gb/s, respectively, after the overall 28% overhead removal. Digital pre-emphasis is used to compensate for the transmitter limited bandwidth. A 4×4 real valued multiple-input-multiple-output (MIMO) adaptive equalizer is used at the receiver for simultaneous polarization tracking and transceiver distortion mitigation instead of a 2×2 complex MIMO filter used in the first demonstration.

The remainder of the chapter is organized as follows: Section 4.2 is dedicated to the DP-IQM description. The experimental setups are discussed in section 3.3. In section 3.4, details on the used DSP functions are presented. Section 3.5 presents the experimental results, and we conclude in section 2.7.

4.2 Description of the deployed InP-based DP-IQM

InP-based modulators offer a significant size reduction compared to LiNbO_3 modulators due to larger phase-voltage efficiency per unit length attributed to the Quantum Confined Stark Effect. The DP-IQM used in this experiment is packaged in a 41 mm by 19 mm module and uses two InP chip-on-carriers as described in [64]. Fig. 4.1 shows a schematic illustrating the DP-IQM design. The input optical power is collimated and split between the two InP chips using a free-space micro-optic 3 dB splitter. A standard aspheric lens is used to focus the light into the InP waveguide and a spot size converter integrated into the modulator was designed to reduce the coupling losses at the waveguide interface. Each InP chip has two nested MZ structures having co-planar slot-line, series push-pull travelling-wave electrodes designed for high-speed RF modulation. Phase bias and substrate bias controls are also required to set operating points and maintain constant modulation voltage over wavelength, respectively. The optical modulated

signal coming out of the Y-polarization chip is rotated by 90° before being combined with the X-polarization signal using a micro-optic polarization combiner. The X-polarization path shows an optical delay component necessary to equalize the optical path between both polarizations to limit the skew to less than 5 ps. The combination of the X and Y signals is collimated and focused inside a standard SMF fiber at the output.

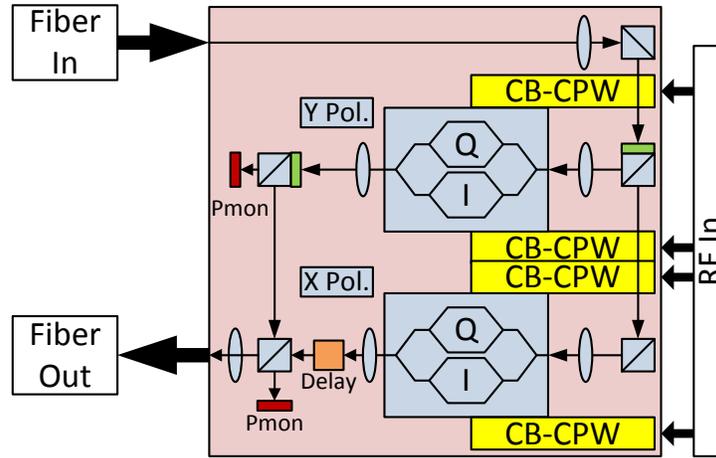


Fig. 4.1 InP-based DP-IQM schematic (CB-CPW: Conduction-backed coplanar waveguide).

The modulator exhibits an insertion loss below 10 dB and a X/Y imbalance of better than 1 dB over the C-band. The V_π for this design is 2.5V and the 3 dB bandwidth is greater than 35 GHz for all the MZ structures.

4.3 Experimental setup

4.3.1 Experimental setup for 400G coherent transmission demonstration

Fig. 4.2 shows the experimental setup deployed to demonstrate the performance of the DP-IQM. An AC-coupled 8-bit digital-to-analog converter (DAC), operated at 65.7 GSps, generates four $1.2 V_{pp}$ differential outputs at the maximum output voltage swing. The four differential outputs are connected to four differential inputs of a quad linear amplifier which produces four $5 V_{pp}$ single-ended signals used to drive the DP-IQM. The optical source is a 100 kHz linewidth external cavity laser (ECL) operating at 1550.12 nm with 15.5 dBm optical power.

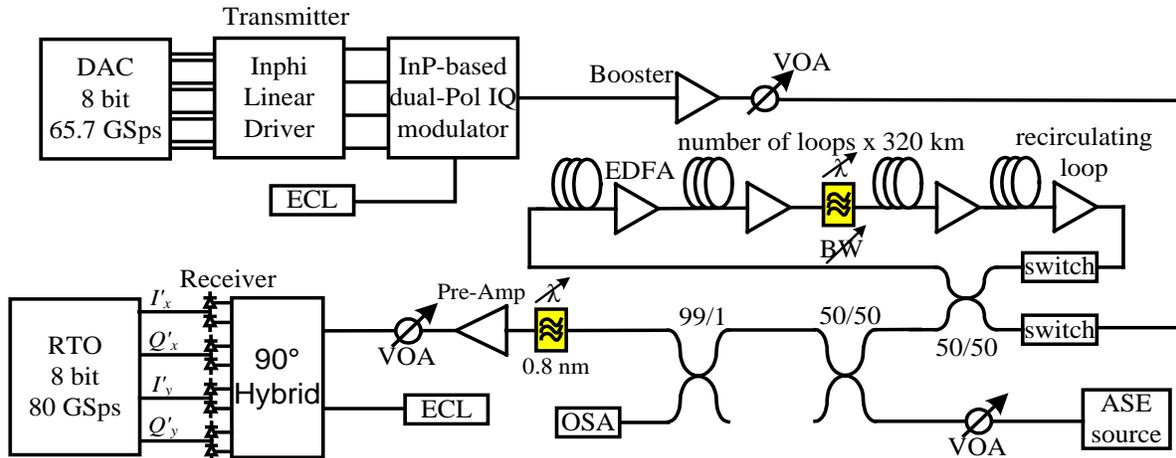


Fig. 4.2 Experimental setup used for 400G coherent transmission demonstration.

After modulation, the optical signal is amplified with a booster to a fixed average power of approximately 23 dBm. The booster is followed by a variable optical attenuator (VOA) used to adjust the launched power of the signal to its optimum value according to the modulation format and symbol rate used. The signal is then launched into an optical recirculating loop containing four spans. Each span contains 80 km of SMF-28e+ followed by an inline erbium-doped fiber amplifier (EDFA) with 5 dB noise figure. The second span is followed by a tunable filter with the center wavelength at 1550.12 nm and bandwidth of 2 nm.

The output signal from the recirculating loop is coupled with a noise signal using a 3-dB coupler for characterization purposes. The noise loading is done using an amplified spontaneous emission (ASE) noise source followed by a VOA to control the OSNR. The output of the 3-dB coupler is followed by a 99% coupler with the 1% port connected to an optical spectrum analyzer (OSA) for OSNR measurements. The 99% port is connected to a 0.8 nm filter followed by pre-amplifier and a second VOA. The optimum received power determined experimentally is 7.4 dBm. This high input power is required in order to compensate for the 10 dB loss of the optical hybrid, thus ensuring adequate received optical power. A 15.5 dBm, 100 kHz linewidth ECL operating at 1550.12 nm is used as a local oscillator laser. The optical hybrid is followed by four balanced detectors feeding a 33 GHz, 8-bit RTO operating at 80 GSps for offline signal processing.

4.3.2 Experimental setup for 600G coherent transmission demonstration

In this subsection, we detail the differences between the 400G demonstration and the 600G demonstration test-beds at the transmitter and the receiver sides. The recirculating loop and the noise loading blocks are kept the same, as shown in Fig. 4.3. Four AC-coupled 8-bit DACs operating at 84 GSps are used to generate the four RF electrical signals, which are the base RF driving signals for optical modulation. The four RF electrical signals are then amplified by four discrete RF amplifiers with single-ended input and output, 50 GHz 3-dB bandwidth and a constant power gain of 26 dB with maximum output voltage of 8 V_{pp}. The actual output swing from the RF amplifiers is around 1.9 V_{pp} at 84 Gbaud and 2.2 V_{pp} at 77 Gbaud. Next, RF delay lines (DC-63 GHz phase adjustable adaptors) are used to compensate for the skew between the four RF signals at the input of the DP-IQM due to RF cabling and other RF components. The DP-IQM is used for electro-optic conversion where the electric fields of two orthogonal polarizations of the continuous wave (CW) input light are optically modulated with the four RF signals. The CW light source is an external cavity laser (ECL) with 100 kHz linewidth operating at 1550.12 nm with an optical power of 15.5 dBm. The DP-IQM is a 2.5V V_π InP DP-IQM which has a 35 GHz 3-dB bandwidth, 10 dB insertion loss and less than 5 ps skew between X and Y polarization paths as detailed in section 4.2.

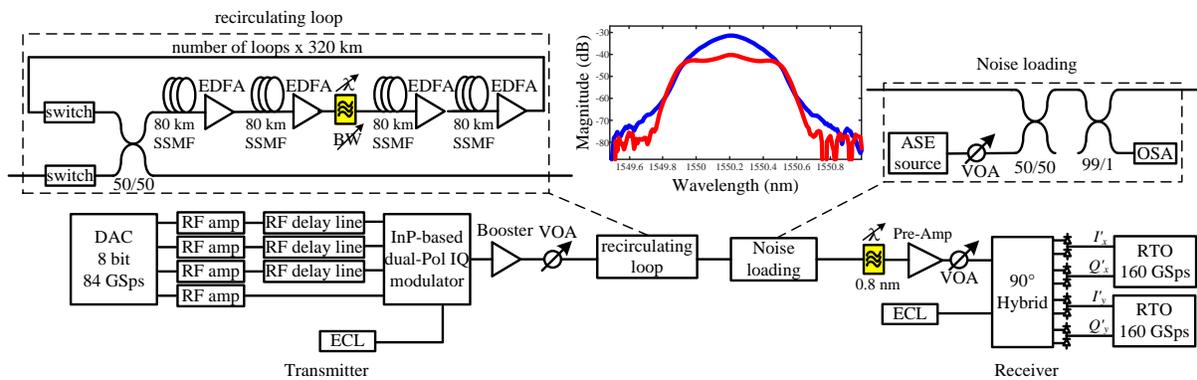


Fig. 4.3 Experimental setup used for 600G coherent transmission demonstration. Inset: B2B optical spectrum with 0.05 nm resolution bandwidth for 84 Gbaud signal with digital pre-emphasis (red curve) and without digital pre-emphasis (blue curve).

For the receiver side, the output of the noise loading block is connected to an 0.8 nm filter to remove the out of band noise followed by pre-amplifier and another VOA to optimize the

received power. The optimized received power is found to be 6 dBm at the signal input port of the optical hybrid. Another ECL operating at 1550.12 nm is used as a local oscillator feeding the optical hybrid for coherent detection. The optical hybrid is followed by four balanced detectors with 40 GHz 3-dB bandwidth which are feeding two 62 GHz, 8-bit, RTOs operating at 160 GSps followed by offline signal processing.

4.4 DSP stack

4.4.1 DSP stack used for 400G coherent transmission

Fig. 4.4 illustrates the DSP stack used at the transmitter and receiver. The transmitter DSP (Tx-DSP) starts with N-QAM symbol generation, generating four random streams for dual polarization transmission. Each stream is uniformly distributed over M levels where $M = \log_2(N)$. Next, the generated symbols are pulse shaped using a root-raised-cosine (RRC) filter at 2 samples per symbol. The roll-off factors used in the RRC filter for 56, 43, 38 Gbaud signals are 0.1, 0.4, and 0.5, respectively. Next, the four streams are re-sampled to the 65.7 GSps sampling rate of the DAC. This is followed by clipping and non-linear compensation for the MZM transfer function, which is needed in order to have equally-spaced constellation points after optical modulation. Then, digital pre-emphasis is used to compensate for the frequency response of transmitter components including (i) the DAC (3-dB bandwidth of 13 GHz), (ii) the quad linear driver (3-dB bandwidth of 30 GHz), and (iii) the DP-IQM (3-dB bandwidth of 35 GHz). Compensation of the overall frequency roll-off of the transmitter minimizes the RF swing out of the DAC. This results in a lower RF signal swing applied to the DP-IQM via the quad linear driver. Therefore, there is a trade-off between the amount of transmitter frequency response compensation applied at the transmitter versus the RF input swing applied to the DP-IQM. The transmitter equalizing filter is experimentally optimized at each symbol rate. The transmitter frequency response compensation is followed by signal quantization to 256 levels for the 8-bit DAC.

The receiver DSP (Rx-DSP) starts with IQ power imbalance compensation and quadrature error correction using Gram-Schmidt orthogonalization [75]. Next, the data is re-sampled from 80 GSps to double the symbol rate. After re-sampling, frequency domain chromatic dispersion (CD) compensation is performed [76] followed by frequency offset removal using the periodogram method [67]. Then, matched filtering is done followed by a timing recovery to

select the appropriate sampling instance [68]. Clock recovery is not required because a common reference clock is supplied to the DAC and RTO. Next, synchronization is performed using a cross-correlation between the received symbols captured by the RTO and the training symbols. Initial polarization tracking is done using the training symbol least mean squares (TS-LMS) algorithm [71]. After initial convergence, decision directed LMS (DD-LMS) for steady-state operation is used. Finally, phase noise and residual frequency offset are mitigated by a second order phase locked loop (PLL) [113] that is included within the TS- and DD-LMS algorithms.

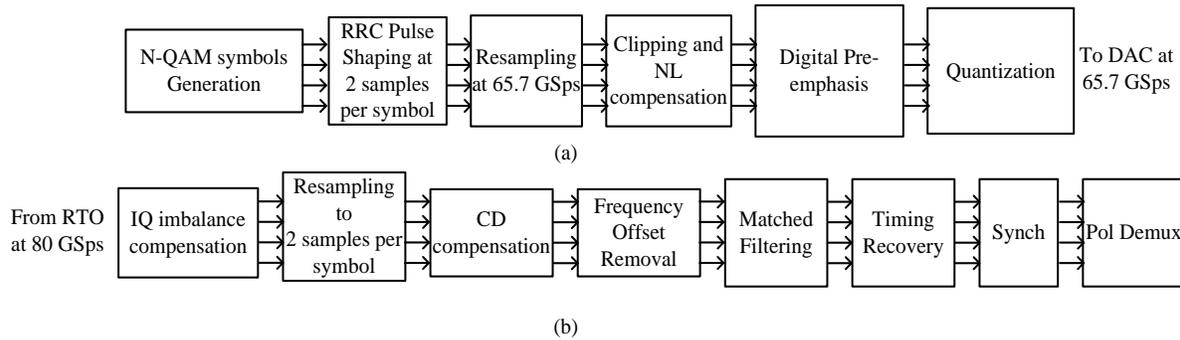


Fig. 4.4 DSP stack used for 400G coherent transmission demonstration: (a) Tx-DSP; (b) Rx-DSP.

4.4.2 DSP stack used for 600G coherent transmission

In this subsection, we highlight differences in the Tx-DSP and Rx-DSP between the 400G demonstration and the 600G demonstration DSP stacks. As shown in Fig. 4.5, the Tx-DSP depends on the operating symbol rate after symbol generation. At 84 Gbaud, which corresponds to operating the DAC at 1 sample per symbol, the generated symbols are pre-distorted by digital pre-emphasis followed by clipping and quantization processes. At 77 Gbaud, the Tx-DSP is the same as in subsection 4.4.1. The generated symbols are initially pulse shaped using RRC filter at 2 samples per symbol with a roll-off factor of 0.09. Then re-sampling from 154 GSps to 84 GSps and digital pre-emphasis are applied on the data samples for equalization. Digital pre-emphasis is achieved using a finite-impulse-response (FIR) filter that is optimized experimentally to give minimum BER in the B2B case. The used number of taps for the FIR filter is 81 taps; however, 99% of the filter energy is confined in 15 taps only. Next, the equalized samples undergo non-linear compensation process to pre-invert the nonlinear transfer function of the MZMs of the DP-IQM and clipping procedure to minimize the peak-to-average-power-ratio. Finally, the samples are quantized and uploaded to the DAC memory.

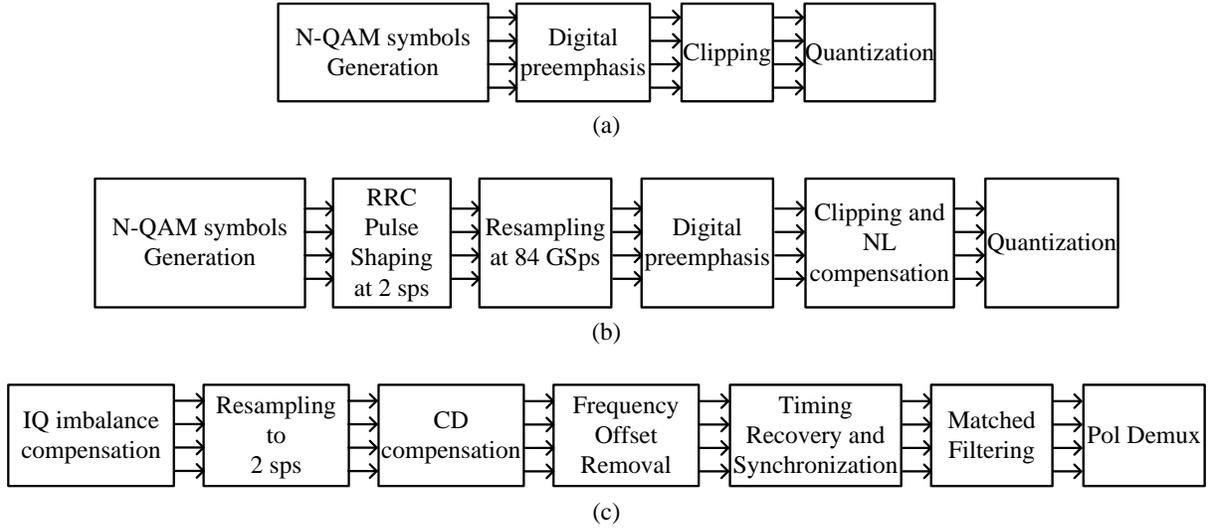


Fig. 4.5 DSP stack used for 600G coherent transmission demonstration: (a) Tx-DSP for 84 Gbaud at one sample per symbol (sps); (b) Tx-DSP for 77 Gbaud; (c) Rx-DSP.

The receiver DSP (Rx-DSP) is similar to the one in subsection 4.4.1 except for timing recovery and polarization demultiplexing blocks. The timing recovery is done during the synchronization process for the captured data from both RTOs. For the polarization demultiplexing block, training symbol least mean squares (TS-LMS) and decision directed LMS (DD-LMS) are used for initial polarization tracking and steady state operation, respectively. However, we use a 4×4 real valued MIMO, similar to [66, 126], instead of the conventional 2×2 complex MIMO for polarization tracking and adaptive filtering. The 2×2 complex MIMO can be described employing Jones space representation of the signal as follows:

$$\begin{bmatrix} \mathbf{E}_X^{out} \\ \mathbf{E}_Y^{out} \end{bmatrix} = \begin{bmatrix} \mathbf{h}_{xx} & \mathbf{h}_{xy} \\ \mathbf{h}_{yx} & \mathbf{h}_{yy} \end{bmatrix} \begin{bmatrix} \mathbf{E}_X^{in} \\ \mathbf{E}_Y^{in} \end{bmatrix} \quad (4.1)$$

where \mathbf{E}_X^{out} and \mathbf{E}_Y^{out} are the recovered complex electric fields after polarization tracking and adaptive filtering of the received complex electric fields \mathbf{E}_X^{in} and \mathbf{E}_Y^{in} using the filter taps \mathbf{h}_{xx} , \mathbf{h}_{xy} , \mathbf{h}_{yy} , and \mathbf{h}_{yx} . Similarly, the 4×4 real valued MIMO can be described by decomposing Eq. (1) into its real valued form and adjusting the real and imaginary parts of the filter taps to be independent as follows:

$$\begin{bmatrix} \mathbf{E}_{X,I}^{out} \\ \mathbf{E}_{X,Q}^{out} \\ \mathbf{E}_{Y,I}^{out} \\ \mathbf{E}_{Y,Q}^{out} \end{bmatrix} = \begin{bmatrix} \mathbf{h}_{xx,ii} & \mathbf{h}_{xx,iq} & \mathbf{h}_{xy,ii} & \mathbf{h}_{xy,iq} \\ \mathbf{h}_{xx,qi} & \mathbf{h}_{xx,qq} & \mathbf{h}_{xy,qi} & \mathbf{h}_{xy,qq} \\ \mathbf{h}_{yx,ii} & \mathbf{h}_{yx,iq} & \mathbf{h}_{yy,ii} & \mathbf{h}_{yy,iq} \\ \mathbf{h}_{yx,qi} & \mathbf{h}_{yx,qq} & \mathbf{h}_{yy,qi} & \mathbf{h}_{yy,qq} \end{bmatrix} \begin{bmatrix} \mathbf{E}_{X,I}^{in} \\ \mathbf{E}_{X,Q}^{in} \\ \mathbf{E}_{Y,I}^{in} \\ \mathbf{E}_{Y,Q}^{in} \end{bmatrix} \quad (4.2)$$

where, $\mathbf{h}_{\theta\theta}$ has been decomposed to $\mathbf{h}_{\theta\theta,ii}$, $\mathbf{h}_{\theta\theta,iq}$, $\mathbf{h}_{\theta\theta,qq}$, and $\mathbf{h}_{\theta\theta,qi}$ and $\theta\theta$ can be any of $\mathbf{x}\mathbf{x}$, $\mathbf{x}\mathbf{y}$, $\mathbf{y}\mathbf{x}$, and $\mathbf{y}\mathbf{y}$. The computational complexity required for multiplications and additions of the real valued MIMO scheme is the same as the complex MIMO, however, it requires more memory for storing the additional independent filter taps.

The advantage of using the 4×4 real valued MIMO over the 2×2 complex MIMO in our setup is its higher ability to mitigate transmitter and receiver impairments namely: IQ imbalance and IQ skew, which are significant impairments for high symbol rates with high order modulation formats. This ability is due to filtering both signal quadratures independently in case of the 4×4 real valued MIMO, which allows the MIMO to correct any time skew or power imbalances between the two quadratures. It is worth noting that the 4×4 real valued MIMO was discussed before for compensating the receiver imperfections in [66, 126]. However, it can be used for transmitter impairment compensation especially for high symbol rate systems that are sensitive to any residual IQ skew or imbalance. In [127], the authors investigated the impact of the transmitter impairments on system performance and introduced a DSP block to compensate for those impairments. This DSP block included a 2×2 real-valued MIMO applied on each polarization separately where polarization tracking is done prior to the transmitter impairments compensation block. In our DSP stack, polarization tracking, carrier phase recovery and the IQ imbalance and IQ skew compensation are combined in one DSP block: the 4×4 real valued MIMO.

4.5 Results and discussion

4.5.1 400G coherent transmission demonstration results

Fig. 4.6 shows the BER versus transmission distance for a 56 Gbaud 16QAM signal at different launched powers. After sweeping the launched power, we found that the optimum launched power is 3 dBm for all transmission distances. At the optimum launched power, 448 Gb/s can be transmitted over 500 km while operating below the HD-FEC threshold. In addition, at 0 km, which is a back-to-back (B2B) configuration with no noise loading, we measure a non-zero BER floor of 2.4×10^{-4} .

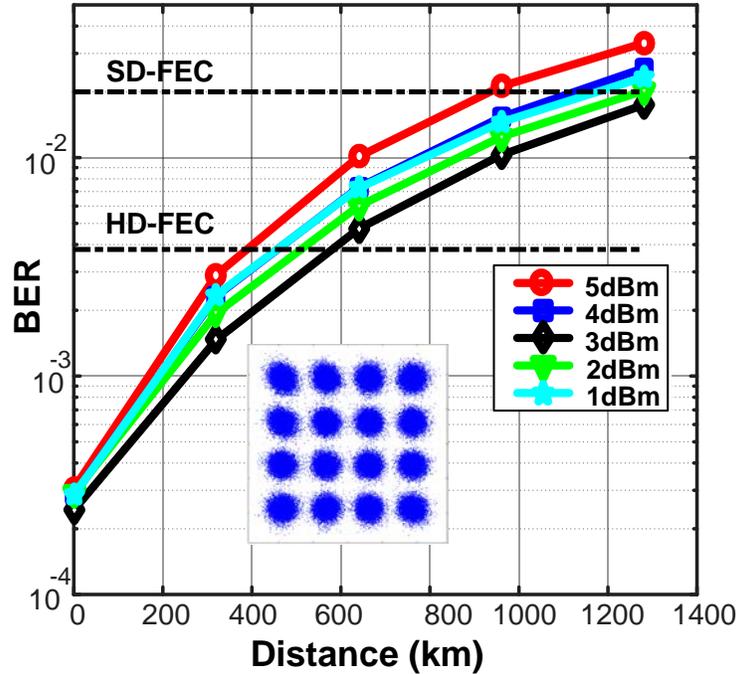


Fig. 4.6 BER versus distance at 56 Gbaud 16QAM per launched power in 400G coherent transmission demonstration with B2B constellation as an inset

Next, Fig. 4.7 shows the BER versus OSNR in 0.1 nm bandwidth for the 56 Gbaud 16QAM signal measured using noise loading in B2B configuration. In addition, we plot the theoretical BER vs. OSNR curve, which assumes white Gaussian noise is added on an otherwise noise free undistorted optical signal. We also include the measured B2B BER floor of 2.4×10^{-4} at an optimum launch power of 3 dBm taken from Fig. 4.6. Because of the presence of this BER floor, we conclude that the experimental OSNR curve at large OSNR values ($\sim > 40$ dB) will converge to this 2.4×10^{-4} BER value. Furthermore, at these high OSNR values the system is operating with very low optical noise because there is no optical noise loading in this configuration. Hence, we conclude that this BER floor is caused by the electrical noise in our system, and that the dominant contribution to this noise is the in-band transmitter electrical noise power. The result is the observed 5 dB implementation penalty between measured and theory curves at HD-FEC. This measured implementation penalty is consistent with recent reports of 16QAM systems operating at similar symbol rates [128, 129].

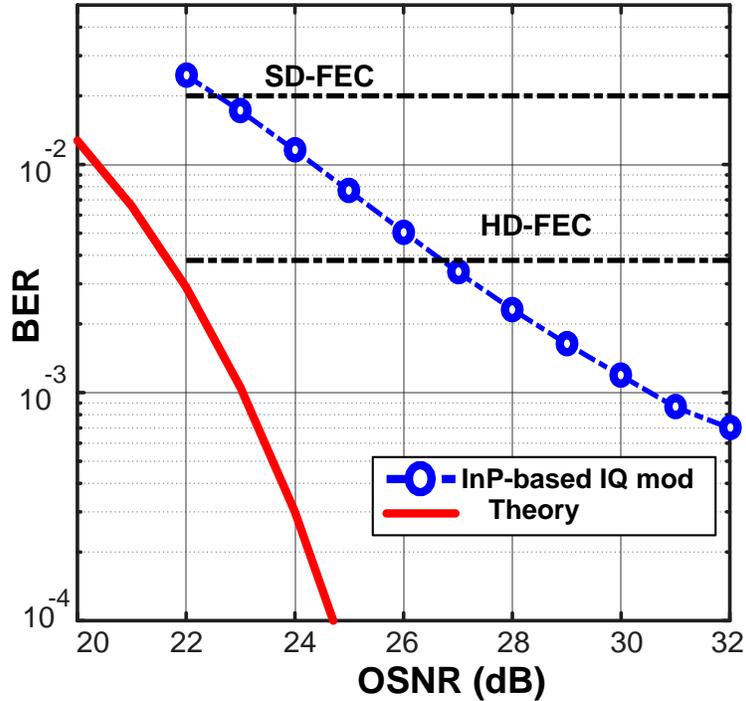


Fig. 4.7 BER versus OSNR for 56 Gbaud 16QAM in 400G coherent transmission demonstration

Regarding the origins of the transmitter electrical noise, the dominant contributions to this electrical noise lying within the signal bandwidth are the DAC, and the RF amplifiers. Contributions to the noise generated by the DAC include: digital waveform clipping, quantization noise, and decreasing effective number of bits (ENOB) with increasing frequency. Further, the RF amplifiers generate thermal noise. In addition generating the IQM drive signals requires pre-emphasis DSP at the transmitter in order to pre-compensate the linear system responses of the components which have limited bandwidths [129, 130]. The resulting signal swing from the DAC post DSP is reduced significantly. Hence, the separation between the four levels used to generate 16QAM signals is very small. This separation will be further reduced when generating eight level signals used in 64QAM formats. This low amplitude signal, which includes excess electrical noise as per above, is amplified by the high gain RF amplifier which adds additional noise. Finally, the large bandwidth of a 56 Gbaud signal integrates large amounts of in-band electrical noise.

Next, Fig. 4.8 shows the BER versus transmission distance at optimum launched power of 2 dBm for both 38 Gbaud 64QAM and 43 Gbaud 64QAM signals both delivering 400G with HD-FEC and SD-FEC overheads, respectively. We observe that using 64QAM modulation format

limits the transmission distance compared to 16QAM. Using 38 Gbaud 64QAM, the maximum reach is limited to 150 km at HD-FEC. The shorter reach for 64QAM compared to 16QAM can be explained by comparing the required OSNR in B2B at HD-FEC in Fig. 4.9 and Fig. 4.7 for both 38 Gbaud 64QAM and 56 Gbaud 16QAM, respectively. We observe that an OSNR of around 32 dB is required at HD-FEC for 38 Gbaud 64QAM (Fig. 4.9) as opposed to only 27 dB for 56 Gbaud 16QAM (Fig. 4.7).

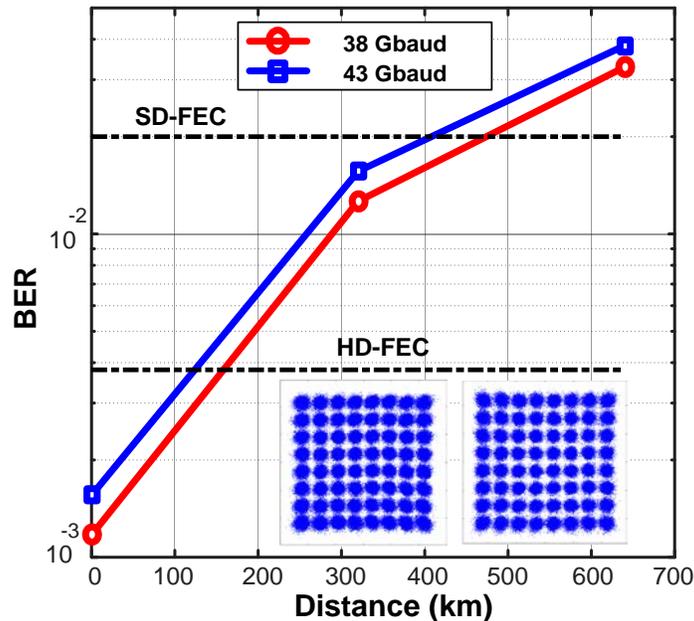


Fig. 4.8 BER versus distance for 38 Gbaud and 43 Gbaud 64QAM in 400G coherent transmission demonstration with B2B constellations as an inset.

Finally, in Fig. 4.9, the B2B BER versus OSNR system performance while operating at 38 Gbaud 64QAM is reported for a 35 GHz InP and a 27 GHz LiNbO₃ DP-IQM. This head-to-head comparison between the two DP-IQMs is performed by removing the InP-based DP-IQM in Fig. 4.2 and installing LiNbO₃ DP-IQM and then optimizing the pre-emphasis filter. We choose this symbol rate, which ensures 400G raw data rate, in order to make a proper comparison between the two platforms such that the impact of the difference in the bandwidth between the two modulators on the system performance is minimized (we use an RRC pulse at 38 Gbaud with a roll-off factor 0.5 which ensures that approximately 91% of the signal energy is confined below 19 GHz). We observe in Fig. 4.9 a difference of approximately 0.3 dB between the performance of the InP modulator and the LiNbO₃ modulator in favor of the InP modulator. In addition, we

observe ~6 dB implementation penalty relative to theory for both modulators, the explanation for which is given above when detailing the 56 Gbaud results.

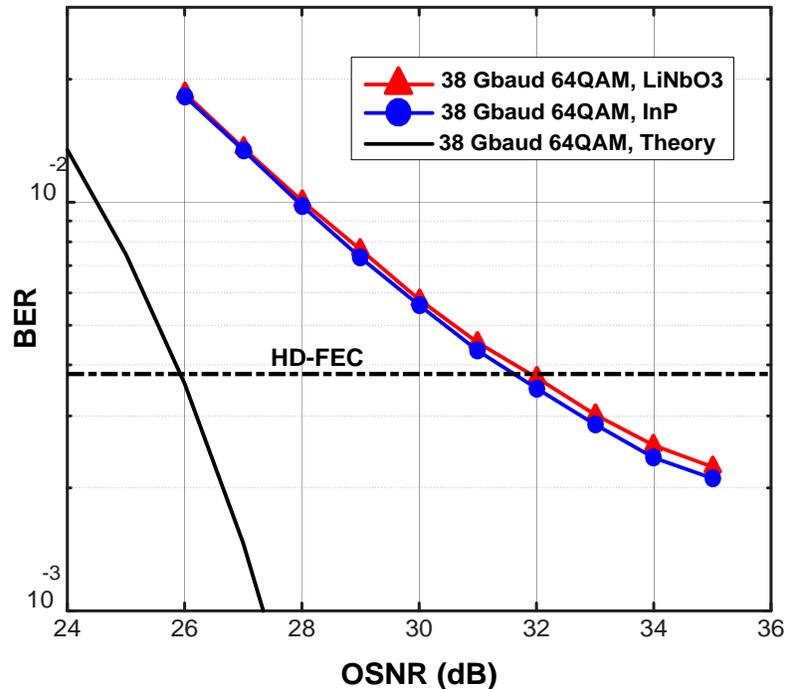


Fig. 4.9 BER versus OSNR for 38 Gbaud 64QAM system performance as a function of modulator platform.

4.5.2 600G coherent transmission demonstration results

As mentioned in subsection 4.4.2, we used 4×4 real valued MIMO instead of the conventional 2×2 complex MIMO to further improve the system performance at these very high symbol rates, which is very sensitive to transceiver I/Q impairments. Fig. 4.10 shows the BER versus transmission distance using the conventional 2×2 complex MIMO and the 4×4 real valued MIMO for 77 Gbaud PDM-32QAM and 84 Gbaud PDM-16QAM, PDM-8QAM and PDM-QPSK signals at optimum launched power of ~3 dBm and received power of 6 dBm. In all cases, the 4×4 real valued MIMO provides better performance than the conventional 2×2 complex MIMO due to being more tolerant to residual IQ imbalance and IQ skew of the transceiver. These transceiver impairments can be attributed to the use of discrete electronic components between the DAC and the DP-IQM. For instance, in the 84 Gbaud 16QAM case, the maximum reach at SD-FEC threshold using 2×2 complex MIMO is 640 km, while the maximum reach

using 4×4 real valued MIMO is 960 km. As per Fig. 4.10, 77 Gbaud PDM-32QAM, 84 Gbaud PDM-16QAM and 84 Gbaud PDM-8QAM signals are propagated and operated below SD-FEC threshold at maximum distances of 320 km, 960 km and 1980 km, respectively. In addition, the 84 Gbaud PDM-QPSK signal is successfully transmitted over 2880 km, operating below the hard decision FEC (HD-FEC) threshold of 3.8×10^{-3} .

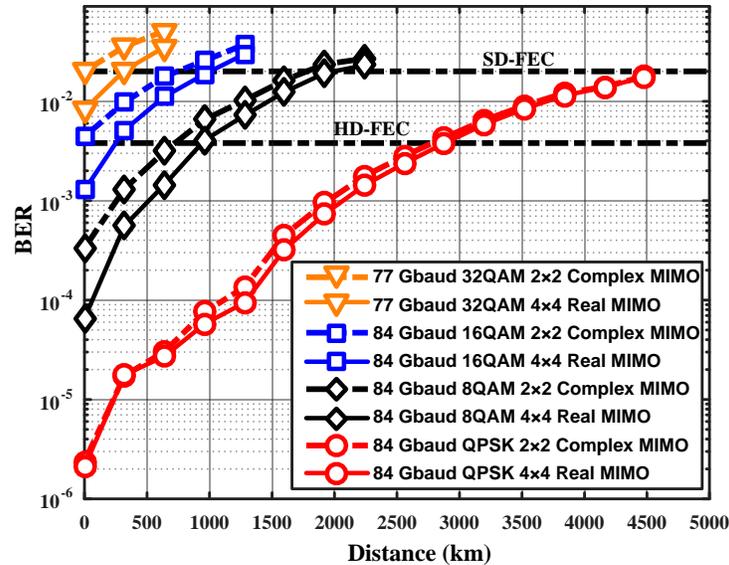


Fig. 4.10 BER versus distance for 77 Gbaud PDM-32QAM signal, 84 Gbaud PDM-16QAM, PDM-8QAM and PDM-QPSK signals for two cases; using 2×2 complex MIMO in Rx-DSP and using 4×4 real valued MIMO in Rx-DSP

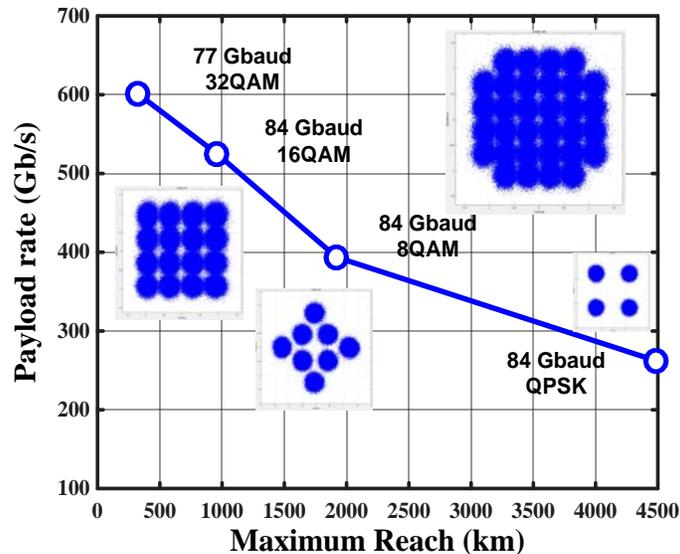


Fig. 4.11 Payload rate after overall overhead removal versus achieved reach (Note: overall overhead 28% in case of SD-FEC). Inset: B2B constellations.

The achieved payload rate after overhead removal versus distance is deduced in Fig. 4.11. A payload rate of 600 Gb/s single carrier transmission over 320 km of SMF is achieved using 77 Gbaud PDM-32QAM modulation format at SD-FEC threshold. Payload rates of 525 Gb/s and 390 Gb/s signals are transmitted at 84 Gbaud over 960 km and 1920 km using 16QAM and 8QAM modulation formats, respectively, at SD-FEC threshold. For 84 Gbaud symbol rate QPSK modulation format, long haul transmission is achieved for payload rates of 300 Gb/s and 260 Gb/s over 2880 km and 4480 km at HD-FEC and SD-FEC thresholds, respectively.

Next, Fig. 4.12 shows the BER versus OSNR in 0.1 nm bandwidth measured by noise loading in B2B configuration for the 77 Gbaud 32QAM, 84 Gbaud 16QAM and 84 Gbaud QPSK signals using the 4×4 real valued MIMO in the Rx-DSP. In addition, we plot the theoretical BER vs. OSNR curves corresponding to these symbol rates and modulation formats assuming additive white Gaussian noise (AWGN) channel. We also include in dotted lines the corresponding measured B2B BER floors taken from Fig. 4.10, i.e. when there is zero loaded ASE noise. Because of the presence of this noise floor, we conclude that by increasing the OSNR, the experimental BER-OSNR curves will converge to the B2B BER values that were achieved in absence of noise loading. These B2B BER floor values result mainly from the in-band electrical transmitter noise as discussed in subsection 4.5.1 and reported in [123, 130].

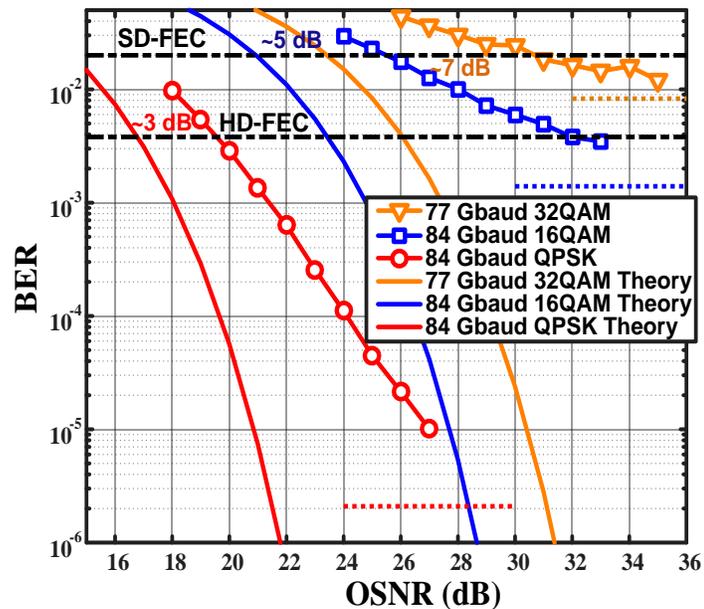


Fig. 4.12 BER versus OSNR for 77 Gbaud 32QAM signal, 84 Gbaud 16QAM and QPSK signals. Dotted lines represent B2B BER floors for each scenario which are taken from the BER curves at 0 km.

It can be noted that the 77 Gbaud 32QAM signal has ~7 dB implementation penalty at SD-FEC threshold, which is higher than the implementation penalty (~5 dB) of the 84 Gbaud 16QAM at SD-FEC threshold despite the fact that the latter has higher symbol rate. This is because the 32QAM signal requires more levels to represent the in-phase and quadrature electrical signals that are generated by the DACs compared to the 16QAM signal. More specifically, higher order modulation formats require higher ENOB for the DACs since they are more sensitive to quantization noise due to the larger number of required levels to represent the underlying drive signals. The same observation can be made for the 84 Gbaud QPSK and 16QAM signals at HD-FEC threshold. The QPSK signal, which has the least ENOB requirement, has ~3 dB implementation penalty at HD-FEC threshold while the implementation penalty for the 16QAM signal is ~8.5 dB at HD-FEC.

4.6 Conclusion

We demonstrated 400G transmission using a 65.7 GSps DAC and an InP-based DP-IQM on a single carrier including the required overhead at different modulation formats and symbol rates, namely 16QAM at 56 Gbaud and 64QAM at both 38 and 43 Gbaud. Transmission of the 56 Gbaud 16QAM signal needs lower OSNR and can propagate over longer distances (more than 500 km) compared to 64QAM signals at lower symbol rates.

Moreover, we demonstrated 770 Gb/s transmission using a 84 GSps DAC and the same InP-based DP-IQM on a single carrier at a symbol rate of 77 Gbaud and PDM-32QAM modulation format over 320 km below SD-FEC threshold. In addition, transmissions are demonstrated at a symbol rate of 84 Gbaud, where bit rates of 504 Gb/s and 672 Gb/s are achieved using 8QAM and 16QAM modulation formats over 960 km and 1920 km, respectively below SD-FEC threshold.

Results in both demonstrations show that the OSNR implementation penalty is highly dependent on the order of the modulation format due to its sensitivity to the transmitter noise. Finally, the viability of using InP DP-IQM is proved to meet up to payload rate of 400 Gb/s and 600 Gb/s transmission system requirements for DCI applications.

Chapter 5

Stokes Vector Direct Detection System for Inter-Datacenter Interconnect Applications

5.1 Overview

Data center interconnects (DCIs) are growing significantly due to the massive increase in data services such as cloud computing, social media and virtualization. This growth is driving the demand for faster, compact and cost-effective optics for DCI traffic in metro applications [131]. Amongst the various solutions, coherent transmission deploying dual-polarization IQ modulators (DP-IQM) and coherent receivers is used for 100G signal transmission with QPSK modulation format in long haul applications. Coherent systems are also used for 200G signal transmission with the 16QAM modulation format for metro applications. These applications have been recently standardized [132, 133]. Metro applications for distances less than 1000 km, found in inter-datacenter interconnects, represent intermediate systems between long haul applications and short reach applications. The transceiver design for metro networks is utmost challenging because of the need of capacity-reach, footprint and cost [131].

Nowadays, there is a growing interest in finding alternative solutions for short reach and metro applications using Stokes vector modulation and detection [78, 79, 81-83, 107, 134, 135] rather than unidimensional intensity modulation/ direct detection (IM/DD) system or complex coherent system. Although the concept of a Stokes vector receiver (SVR) has been described in the 1990s [77, 136], SVR systems with advanced modulation formats have only been demonstrated recently. For instance, 100 Gb/s per wavelength over 480 km for metro applications was demonstrated in [79] using a complex IQ modulation on the X polarization and transmitting the carrier on the orthogonal polarization. This modulation scheme presents a hybrid solution between coherent and direct detect systems.

The two-dimensional Stokes vector direct detection (2D-SV-DD) system and single polarization coherent system have a simpler transmitter when compared to that of the dual polarization coherent system. Since only one polarization is modulated, the transmitter in 2D-SV-DD requires half the number of electronic and electro-optic devices including the number of

digital-to-analog converters (DACs), amplifiers, Mach-Zehnder modulators compared to the dual polarization coherent system transmitter. This is at the expense of halving the spectral efficiency that can be achieved by the dual polarization coherent system. On the other hand, the SVR, which has several configurations shown in [82], has a simpler receiver compared to the single polarization coherent system. The common feature in SVR configurations compared to coherent systems is that SVR does not employ a local oscillator (LO) laser with its associated thermoelectric cooler (TEC) and driving circuitry. The deployed SVR in this demonstration comprises a 2×4 hybrid, two balanced photodetectors (BPDs) and two single-ended photodetectors (SE-PDs) instead of two 2×4 hybrids and 4 BPDs in coherent detection. It is worth noting that the single polarization and the dual polarization coherent receivers share the same configuration to avoid the deployment of a polarization stabilizer in single polarization coherent receiver [73]. Moreover, 2D-SV-DD systems require less complex digital signal processing (DSP) at the receiver compared to a coherent system since there is no frequency offset removal, laser phase noise mitigation or cycle slip detection.

In parallel to the interests in 2D-SV-DD systems, researchers, designers, and manufacturers use photonic integrated devices fabricated on various platforms such as silicon (Si) [39] or indium phosphide (InP) [52] to supply the demand for DCI applications. 2D-SV-DD systems have been demonstrated using Si photonics modulators in [78, 83] and using Si photonics IQ modulator (IQM) and Si photonics SVR in [107].

In this chapter, we first demonstrate, using an InP IQM described in section 4.2, a single carrier transmission using a 2D-SV-DD system at symbol rate of 56 Gbaud with QPSK, 8QAM, and 16QAM modulation formats corresponding to payload rates of 100G, 150G and 200G, respectively [58]. The system operates below the 7% overhead BCH hard-decision forward error correction (HD-FEC) threshold of 4.5×10^{-3} [137] over distances of 2880, 960 and 320 km, respectively. These modulation schemes enable 100G transmissions and 200G transmissions for long haul and metro applications after removal of the FEC and encapsulation overhead. To achieve this result, we optimize the carrier-to-signal-power ratio (CSPR) and implement DSP at the transmitter and the receiver including digital pre-emphasis, non-linear compensation for the Mach-Zehnder modulator (MZM) transfer function, and chromatic dispersion (CD) pre-compensation. The results of this phase of 2D-SV-DD demonstration is enabled by a DAC

operating at 64 GSps in the transmitter side and real time oscilloscope (RTO) with 33 GHz 3-dB bandwidth operating at 80 GSps at the receiver side.

We then investigate the effect of the used CD compensation method on the 2D-SV-DD transmission system performance. We discuss the limitations of both CD pre- and post-compensation methods and we identify the transmission scenarios in which either method outperforms the other as well as the potential reasons for the performance discrepancy between the two compensation methods. In addition, we show mathematically the importance of performing the polarization tracking prior to CD post-compensation. Specifically, we compare the transmission performance of CD pre-compensation and CD post-compensation at symbol rates of 64 Gbaud and 84 Gbaud using 16QAM modulation format enabled by a DAC operating at a sampling rate of 84 GSps. Results show that the CD post-compensation gives better performance than the CD pre-compensation at 84 Gbaud. On the other hand, CD pre-compensation outperforms CD post-compensation when the symbol rate is dialed down to 64 Gbaud. Also, we show the system performance for different modulation formats at 64 Gbaud and 84 Gbaud using CD pre-compensation showing the maximum reach of each format. We show that using 2D-SV-DD combined with CD pre-compensation, the maximum reach can extend to thousands of kilometers at BER below the soft decision forward error correcting threshold (SD-FEC) of 2×10^{-2} especially for the less dense modulation formats. For instance, 84 Gbaud QPSK and 64 Gbaud QPSK transmissions reached 3,520 km and 4,800 km below SD-FEC achieving 591,360 Gb/s.km and 614,400 Gb/s.km throughput-times-distance products, respectively, which are the highest throughput-times-distance products reported using Stokes transmission.

The rest of this chapter is organized as follows: In section 5.2, we present the principle of the 2D-SV-DD system under study and the effect of the polarization rotation due to propagation on CD compensation. The experimental setup is discussed in section 5.3 for the two phases of our study. In section 5.4, the applied DSP functions in both phases are detailed and differences in DSP are highlighted when either CD pre- or post-compensation is used. Section 5.5 presents the experimental results, and we finally conclude the chapter in section 5.6.

5.2 Principle of 2D-SV-DD system and CD limitations

5.2.1 Principle of 2D-SV-DD system transceiver

A schematic diagram of the 2D-SV-DD system is shown in Fig. 5.1. After the transmitter DSP (Tx-DSP) which will be detailed in the next section, the processed samples are loaded to the DAC followed by RF amplifiers to drive a single polarization IQ modulator. The IQ modulator performs complex modulation on the electric field of one of the two orthogonal SOPs originating from a single laser (X-polarization) that is denoted by E_X^t . The power level of optical carrier (C) on the other SOP (Y-polarization) is adjusted for the optimization of the carrier-to-signal power ratio (CSPR). Then, both optical signals (the modulated signal and the optical carrier) are combined using a polarization beam combiner (PBC). The co-propagating reference carrier allows for the retrieving of the modulated signal during the self-beating process at the SVR. It is noted that a SV-DD system shares identical transmitter architecture and almost similar receiver architecture with a self-homodyne coherent detection system that uses a polarization-multiplexed pilot carrier. However, by adding two single-ended photodetectors, SVR does not require a polarization stabilizer to separate the pilot tone and the signal that is essential for a self-homodyne coherent detection system [138, 139].

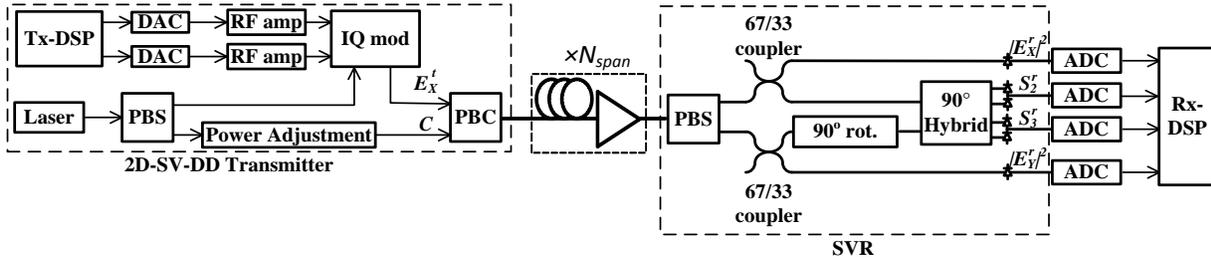


Fig. 5.1 Schematic diagram of the 2D-SV-DD system.

The transmitted field can be equivalently represented in the Jones space $\mathbf{J} = [E_X^t \quad C]^T$ or in the Stokes space $\mathbf{S}^t = [S_0^t \quad S_1^t \quad S_2^t \quad S_3^t]^T$, where both representations can be related by $S_0^t = |E_X^t|^2 + C^2$, $S_1^t = |E_X^t|^2 - C^2$, $S_2^t = 2\text{Re}\{E_X^t \cdot C^*\}$, $S_3^t = 2\text{Im}\{E_X^t \cdot C^*\}$. \mathbf{T} denotes a vector transpose, the superscript ‘*’ stands for the complex conjugate and Re and Im represent the real and imaginary parts of a complex number. Without loss of generality if we assume that $C = 1$ and $E_X^t = I + jQ$, we get $S_0^t = |E_X^t|^2 + 1$, $S_1^t = |E_X^t|^2 - 1$, $S_2^t = 2I$, $S_3^t = 2Q$. This means that

the valuable information is available on S_2^t and S_3^t . After propagation in the fiber, the SOP of transmitted signal is randomly rotated due to fiber birefringence. This rotation is modeled in either the Jones space by an unknown unitary 2×2 complex valued matrix that multiplies the transmitted Jones vector \mathbf{J} or the Stokes space by the matrix multiplication of an unknown 4×4 real valued matrix and the transmitted Stokes vector \mathbf{S}^t . Due to random SOP rotation along the fiber, each of the received electric fields on X and Y polarizations (E_X^r and E_Y^r) is a linear combination of the transmitted polarizations E_X^t and E_Y^t .

The received Stokes vector $\mathbf{S}^r = [S_0^r \ S_1^r \ S_2^r \ S_3^r]^T$ can be measured using the Stokes vector receiver (SVR) illustrated at the right hand side of Fig. 5.1. A PBS is initially used to split the received signal into two orthogonal SOPs. Each SOP is further split using 67/33 polarization maintaining couplers. This coupling ratio ensures that the system performance is independent from the SOP of the signal before the PBS [83]. At the 33% output port of the couplers, $|E_X^r|^2$ and $|E_Y^r|^2$ are detected separately using single-ended photodetectors to obtain the intensity on both polarizations giving the terms required for calculating $S_0^r = |E_X^r|^2 + |E_Y^r|^2$, and $S_1^r = |E_X^r|^2 - |E_Y^r|^2$. At the 67% output ports of the coupler, the two orthogonal SOPs are introduced to a 90° optical hybrid after SOP alignment using a 90° polarization rotator on one branch. After balanced detection at the output of the hybrid, the beating provides the received Stokes parameters $S_2^r = 2\text{Re}\{E_X^r \cdot E_Y^{r*}\}$, $S_3^r = 2\text{Im}\{E_X^r \cdot E_Y^{r*}\}$. Since S_2^t and S_3^t are the only transmitted Stokes parameters that are required to be retrieved, they can be found via the following 3×2 MIMO filter that performs the SOP de-rotation,

$$\begin{bmatrix} S_2^t \\ S_3^t \end{bmatrix} = \begin{bmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \end{bmatrix} \begin{bmatrix} S_1^r \\ S_2^r \\ S_3^r \end{bmatrix} \quad (5.1)$$

where m_{11}, \dots, m_{23} are coefficients of the inverse rotation matrix that are found adaptively at the receiver via DSP, and are used to recover the Stokes parameters S_2^t and S_3^t which in turn fully describe the complex field of the modulated signal.

5.2.2 CD limitations in 2D-SV-DD system

In this subsection, we discuss the CD limitations in this 2D-SV-DD system. Specifically, we show how the CD compensation is performed by analyzing the Jones space and Stokes space representations of the 2D-SV-DD system. After propagation in the fiber, CD distorts the signal

besides the random rotation of the SOP of the propagated signal due to fiber birefringence. If we assume a general rotation around the Poincaré sphere, it can be modeled together with CD in the Jones space as follows:

$$J^r = \begin{bmatrix} E_X^r \\ E_Y^r \end{bmatrix} = \widehat{D} \begin{bmatrix} a & b \\ -b^* & a^* \end{bmatrix} J^t \quad (5.2)$$

where J^r is the received Jones vector, a and b are the complex entries of the general polarization rotation matrix [140] and $\widehat{D} = -j \frac{\beta_2 L}{2} \frac{\partial^2}{\partial t^2}$ is a linear operator representing CD encountered by the signal along fiber of length L and dispersion parameter β_2 . It is clear that CD affects both polarizations of the propagating lightwave similarly. Also, since the carrier C (the Y -component of J^t) represents CW light that is nearly constant with time, it is not largely impacted by CD whereas \widehat{D} only distorts E_X^t leading to intersymbol interference besides polarization rotation. In the absence of polarization rotation Eq. (5.2) is rewritten as

$$J^r = \widehat{D} J^t = \widehat{D} \begin{bmatrix} E_X^t \\ C \end{bmatrix} = \begin{bmatrix} \widehat{D} E_X^t \\ C \end{bmatrix}$$

in which case

$$S_2^r + jS_3^r = \widehat{D} \{2E_X^t \cdot C^*\} = \widehat{D} \{S_2^t + jS_3^t\}$$

Hence in the absence of polarization rotation, CD can be easily post compensated by applying the inverse operator \widehat{D}^{-1} on the received $S_2^r + jS_3^r$ after which $S_2^t + jS_3^t$, i.e., the complex data symbols, are recovered. However in the presence of polarization rotation, applying \widehat{D}^{-1} to $S_2^r + jS_3^r$ does not recover the transmitted data symbols. This can be deduced by writing $S_2^r + jS_3^r$ as follows

$$S_2^r + jS_3^r = 2E_X^r E_Y^{r*} = a^2 \widehat{D} \{S_2^t + jS_3^t\} - b^2 (\widehat{D} \{S_2^t + jS_3^t\})^* + 2ab \left(|\widehat{D} E_X^t|^2 - |C|^2 \right)$$

to which applying \widehat{D}^{-1} clearly does not only inverse the CD, but also adds noisy terms that affect the polarization de-rotation to recover the transmitted data $S_2^t + jS_3^t$. From the above analysis, we conclude that if CD post-compensation is to be used at the receiver, polarization rotation has to be initially corrected at the receiver before \widehat{D}^{-1} is applied. However, if CD pre-compensation is adopted, \widehat{D}^{-1} is simply applied to $S_2^t + jS_3^t$ or E_X^t prior to transmission which is then canceled by the \widehat{D} operator of the CD induced by the propagation along the fiber.

In summary, CD can be either post- or pre-compensated in a single carrier 2D-SV-DD. If CD post-compensation at the receiver is the chosen method, it should be performed after the

polarization tracking to have a correct implementation that is independent of the SOP of the received signal. On the other hand, CD pre-compensation is performed at the transmitter by applying the inverse CD operator on the complex data symbols before propagation. In the following sections, we will clarify how the CD compensation method affects the Tx-DSP and Rx-DSP stacks and how each compensation method performs depending on operating system parameters.

5.3 Experimental setup

5.3.1 First phase of 2D-SV-DD system demonstration

Fig. 5.2 shows the experimental setup deployed to demonstrate the performance of the SV transmission system. An AC-coupled 8-bit digital-to-analog converter (DAC) operating at 64 GSps generates two $1.2 V_{pp}$ differential outputs. These two differential outputs are amplified with a linear amplifier. The linear driver produces two $5 V_{pp}$ single-ended signals that drive one half of the DP-IQM. This represents the complex modulation on the X-Pol of the DP-IQM. The InP DP-IQM has a 35 GHz 3-dB bandwidth, less than 10 dB of insertion loss, a V_{π} of 2.5V, and less than 5 ps skew between both polarization paths [59]. To control the carrier power on Y-pol relative to the fixed modulated signal power on the X-pol, the MZM on the Q component of Y-pol is biased at null and the bias for the MZM on the I component of Y-pol is swept to produce the desired CSPR. The optical source is a 100 kHz linewidth external cavity laser (ECL) operating at 1550.12 nm with 15.5 dBm optical power.

Before transmission, the optical signal is amplified using a booster EDFA to compensate for subsequent switches and couplers loss used in the recirculating loop and to enhance the signal power level for better optical signal to noise ratio during transmission. The optimization of the total launched signal power depends on the optimization of the CSPR and launching sufficient power for the modulated signal on the X polarization according to the used modulation format to have the required OSNR for the achieved reach. This dependency on CSPR value is because of the booster which has a fixed output power of 23 dBm, so the increase in the CSPR value is accompanied by a decrease in the modulated signal power on the X polarization. Thus, the launched signal power is optimized using a variable optical attenuator (VOA) along with the CSPR for different modulation formats by minimizing the bit error rate at the maximum reach according to the used modulation format. The signal is then transmitted into a four-span optical

recirculating loop. Each span comprises 80 km of SMF-28e+ followed by an in-line EDFA with 5 dB noise figure. A tunable optical filter is used after the second span with its center wavelength set to 1550.12 nm and bandwidth set to 2 nm.

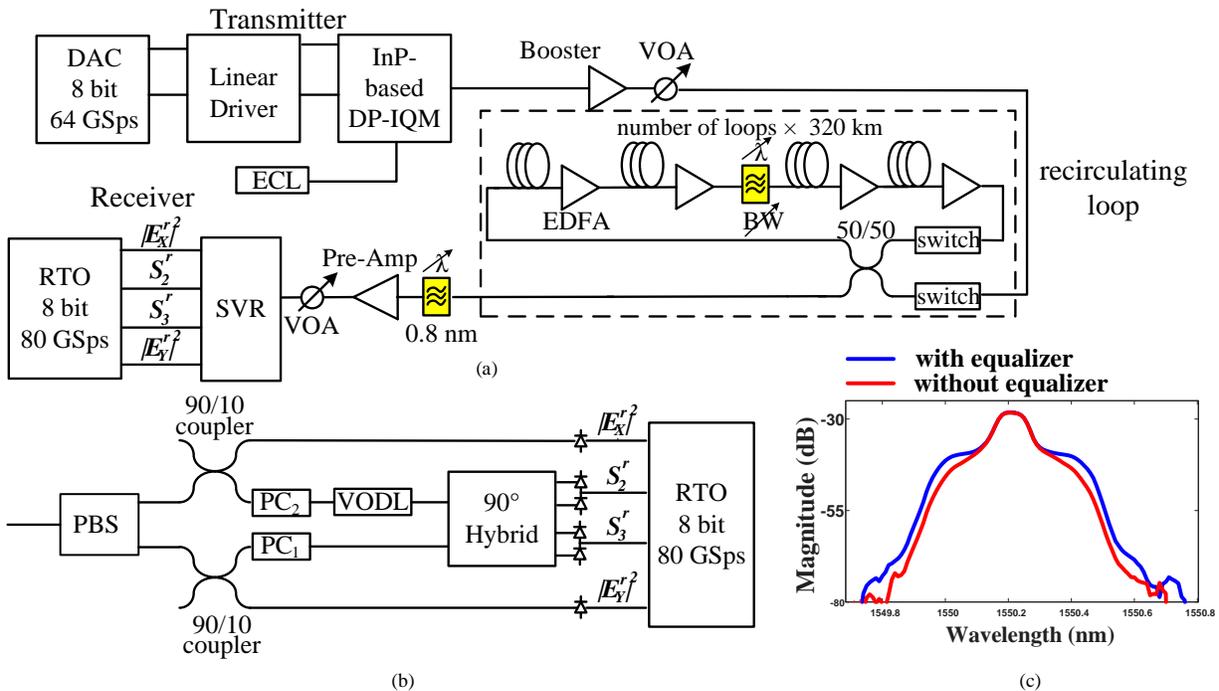


Fig. 5.2 First phase of 2D-SV-DD demonstration (a) Experimental setup (b) SVR details (c) B2B optical spectrum with 0.05 nm resolution bandwidth.

The output signal from the recirculating loop is connected to a 0.8 nm filter followed by a pre-amplifier and a second VOA. The optimum received power, as determined experimentally, is 16.6 dBm. This high required received optical power along with the use of 90/10 coupler instead of 67/33 coupler are needed to compensate for the large insertion loss of all passive discrete optical components that are connected on the optical bench to realize the middle section of an SVR. In particular, the use of a 2×8 dual polarization optical hybrid (which has a theoretical insertion loss of 9 dB) instead of a 2×4 single polarization hybrid to realize an SVR (see Fig. 5.1) accounts for an additional 3 dB insertion loss that could have been avoided if a single polarization device is available for the experiment. To operate only one half, i.e., one polarization, of the dual polarization hybrid, two polarization controllers (PC₁ and PC₂ in Fig. 5.2) at its inputs are adjusted such that the two input fields beat in one half of the device. At the input of the SVR, a PBS initially splits the incoming signal into two received orthogonal

polarizations. Each polarization is then split using a 90/10 polarization maintaining coupler where the 10% port is connected directly to a single-ended photodetector to detect the intensity on each received polarization ($|E_X^r|^2$ and $|E_Y^r|^2$). The 90% ports are connected to the hybrid, with one of its inputs rotated by 90° as shown in Fig. 5.1. A variable optical delay line (VODL) is used to match the optical delay between the two polarization tributaries of the signal before entering the optical hybrid to have correct signal beating. The optical hybrid is followed by balanced detection which provides the received Stokes parameters (S_2^r and S_3^r). The direct detection terms ($|E_X^r|^2$ and $|E_Y^r|^2$) and the Stokes parameters (S_2^r and S_3^r) are then sampled for offline signal processing using a real time oscilloscope (RTO) having four synchronized channels each is running at sampling rate of 80 GSps and having an analog bandwidth of 33 GHz.

5.3.2 Second phase for CD impact on 2D-SV-DD demonstration

In this subsection, we highlight the differences in the second phase experimental setup from that of the first phase. Fig. 5.3 shows the experimental setup deployed for the second phase 2D-SV-DD system demonstration. An AC-coupled four-channel 8-bit digital-to-analog converter (DAC) operating at 84 GSps generates two $0.6 V_{pp}$ single-ended outputs. These two single-ended outputs are amplified with two RF amplifiers with 50 GHz 3-dB bandwidth. One of the RF amplifiers is followed by an RF delay line with 63 GHz 3-dB bandwidth for adjusting the RF skew between the two RF paths. It is worth noting that the skew is divided into two parts; one part is an integer number of samples, which is resolved in the Tx-DSP by simply shifting the samples on one of the DAC channels. The second part, which is a fraction of a sample, is adjusted using the RF delay line to avoid unnecessary effect on the signal quality if the total skew was to be adjusted in the Tx-DSP. After amplification and skew adjustment, these two RF signals are used to drive one half of the InP DP-IQM which represents the complex modulation on the X-polarization of the DP-IQM. The CSPR is controlled by adjusting the bias of the MZMs on the Y-polarization of the DP-IQM. The DP-IQM is fed by external cavity laser (ECL) operating at wavelength of 1550.12 nm with 100 kHz linewidth and 15.5 dBm optical power. After optical modulation, the optical signal is amplified using a booster EDFA with a fixed output power level of 23 dBm followed by a variable optical attenuator (VOA) to optimize the total launched power. Also, the booster EDFA allows for compensating the loss of the optical

recirculating loop front end (switches and couplers). The recirculating loop is the same as the one used in subsection 5.3.1.

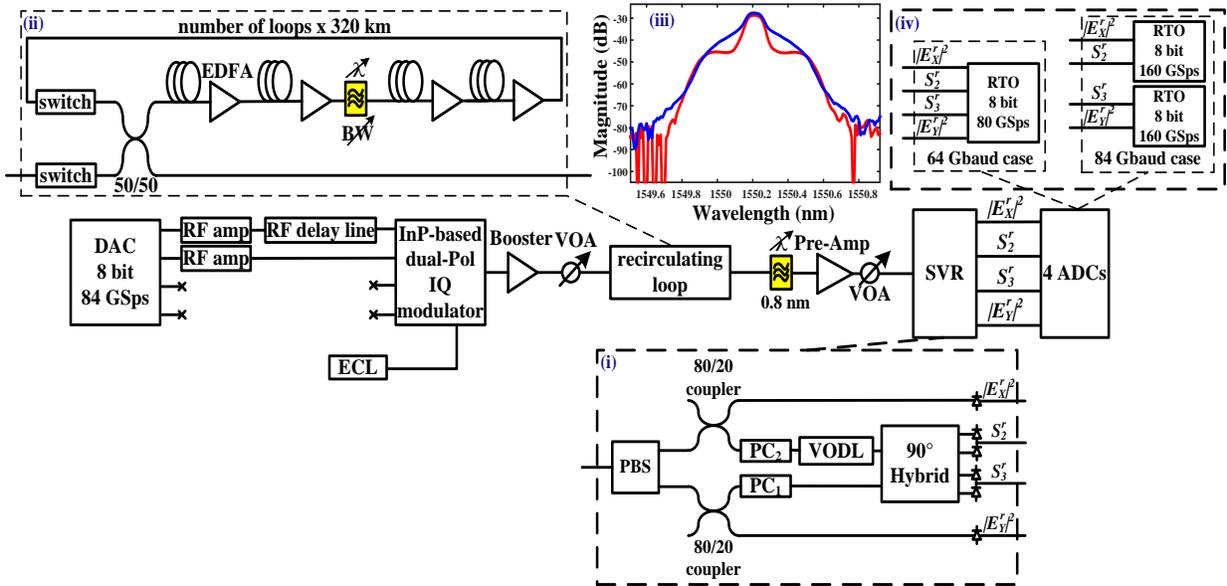


Fig. 5.3 Second phase of 2D-SV-DD demonstration experimental setup showing the following insets: (i) SVR details, (ii) recirculating loop details, (iii) B2B optical spectrum with 0.05 nm OSA resolution bandwidth in case of the 84 Gbaud signal, and (iv) RTOs configuration according to the operating symbol rates.

The output signal from the recirculating loop is filtered using 0.8 nm filter to remove the out-of-band noise. The filter is followed by a pre-amplifier and a second VOA to optimize the received signal power which is determined experimentally to be 16 dBm. The deployed SVR in this subsection is identical to the one used in Fig. 5.2(b) except for using an 80/20 polarization maintaining coupler instead of 90/10 coupler. Since the optical hybrid accounts for the main loss in the SVR, the 80% ports are directed to the middle part of the SVR shown in Fig. 5.3(i) which includes the optical hybrid. The 80% ports of the couplers are connected to two polarization controllers which are adjusted such that the two input fields beat in one half of the hybrid. A variable optical delay line (VODL) is used to match the optical delay between the two paths before entering the optical hybrid to have correct signal beating. The optical hybrid is followed by balanced detection which provides the received Stokes parameters (S_2^r and S_3^r). The 20% ports are connected directly to SE-PDs to detect the direct detection terms (the square of the electric fields) on both orthogonal SOPs ($|E_x^r|^2$ and $|E_y^r|^2$). The direct detection terms ($|E_x^r|^2$ and $|E_y^r|^2$) and the Stokes parameters (S_2 and S_3) are then sampled for offline signal

processing. In the case of the 64 Gbaud signal transmission, a real-time oscilloscope (RTO) is used serving as a four-channel ADC running at sampling rate of 80 GSps with analog bandwidth of 33 GHz as shown in Fig. 5.3(iv). In the case of the 84 Gbaud signal transmission, two RTOs are used for sampling. Each RTO has two real edge channels running at sampling rate of 160 GSps with analog bandwidth of 62 GHz.

5.4 DSP stack

5.4.1 First phase of 2D-SV-DD system DSP stack

Fig. 5.4 illustrates the DSP stack used at the transmitter and receiver. The offline transmitter DSP (Tx-DSP), shown in Fig. 5.4(a), starts with the N-QAM symbol generation. Next, the generated symbols are pulse shaped using a raised-cosine (RC) filter having a roll-off factor of 0.1 at two samples per symbol. Afterwards, the signal is pre-distorted for chromatic dispersion (CD) pre-compensation depending on the propagation distance and fiber type. The chromatic dispersion is fully pre-compensated at the transmitter side which allows combining the polarization tracking and receiver equalizer in one DSP block.

In this 2D-SV-DD system with complex modulation, CD can also be compensated at the receiver. Applying the CD post-compensation can be done perfectly when the state of polarization (SOP) of the input signal is adjusted such that the transmitted carrier is on one input port of the hybrid and the modulated complex signal is on the other port. However, when the SOP is random, severe penalties arise when attempting to do CD post-compensation in the presence of the two power terms $|E_X^t|^2$ and $|C|^2$ appearing in the S_2^r and S_3^r received Stokes parameters. This prevents CD post-compensation before polarization tracking or at least will affect the performance considerably as detailed in the next subsection. Another method is to do CD-post-compensation after the MIMO filter which performs polarization tracking. This method, however, requires the use of slow training symbols to obtain the Stokes vector MIMO de-rotation matrix prior to doing CD post-compensation, as was done in [81]. Details about using different methods of CD compensation and its impact on the DSP stack will be presented in the next subsection.

After CD pre-compensation, the I and Q streams of the N-QAM symbols are re-sampled to 64 GSps to match the sampling rate of the subsequent DAC as shown in Fig. 5.4(a). The inverse of the non-linear transfer function of the MZM is then applied to the resampled symbols such that

the constellation points are equally spaced after optical modulation. Afterwards, an experimentally optimized FIR filter is used in the digital pre-emphasis block to partially equalize for the frequency response of the transmitter components including the DAC, the linear driver, and the InP DP-IQM. Finally, the equalized samples are quantized to 256 levels and uploaded to the memory of the 8-bit DAC.

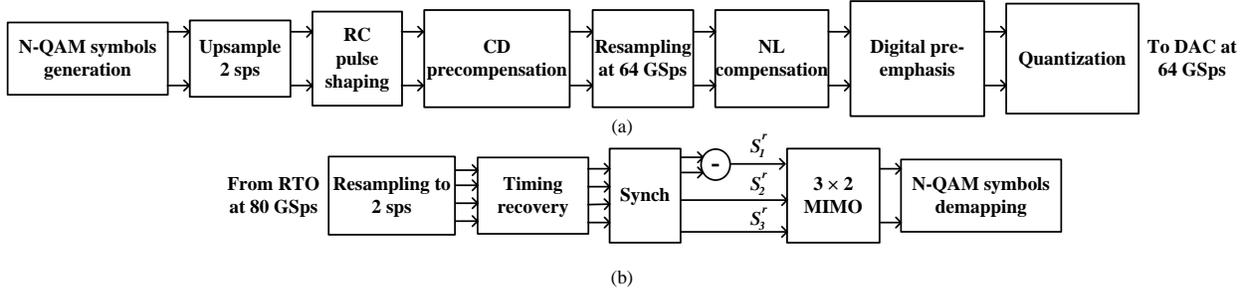


Fig. 5.4 DSP stack used for the first phase of 2D-SV-DD demonstration: (a) Tx-DSP; (b) Rx-DSP.

The offline receiver DSP (Rx-DSP) starts by re-sampling the captured patterns from 80 GSps to twice the symbol rate because all Rx-DSP operates at two samples per symbol. After re-sampling, a timing recovery to select the appropriate sampling instance is performed using the same algorithm as in [59]. After constructing the S_1^r parameter from the two direct detection terms ($|E_X^r|^2$ and $|E_Y^r|^2$), the 3×2 real valued MIMO adaptive filtering is used for two roles: (i) it tracks the polarization rotation and inverts it according to Eq. (5.1) to recover the transmitted Stokes parameters S_2^t and S_3^t and (ii) it equalizes for any residual inter-symbol interference (ISI). In the 3×2 MIMO, training symbol least mean squares (TS-LMS) algorithm is used initially to track the polarization rotation and determine the adaptive filter tap coefficients. After the training symbols, decision directed LMS (DD-LMS) replaces TS-LMS where symbol decisions in the steady-state operation are used to slowly adapt the taps of the 3×2 MIMO. Finally, the received symbols are de-mapped into bits for BER counting. The BER counting is done over 13,762,560 bits, 10,321,920 bits, and 6,881,280 bits, in case of 16QAM, 8QAM and QPSK, respectively.

5.4.2 Second phase DSP stack for CD impact on 2D-SV-DD

Fig. 5.5 illustrates the DSP stack used at the transmitter and receiver in two cases depending on which CD compensation method is adopted: CD pre-compensation shown in Fig. 5.5(a) and CD post-compensation shown in Fig. 5.5(b). The offline transmitter DSP (Tx-DSP) stack in Fig. 5.5(a) depends on the oversampling factor used at the transmitter, i.e., the DAC sampling rate relative to the symbol rate. In our experiment, the DAC sampling rate is fixed at 84 GSps and two different symbol rates are tested: 84 and 64 Gbaud representing oversampling factors of 1 and 1.3125, respectively.

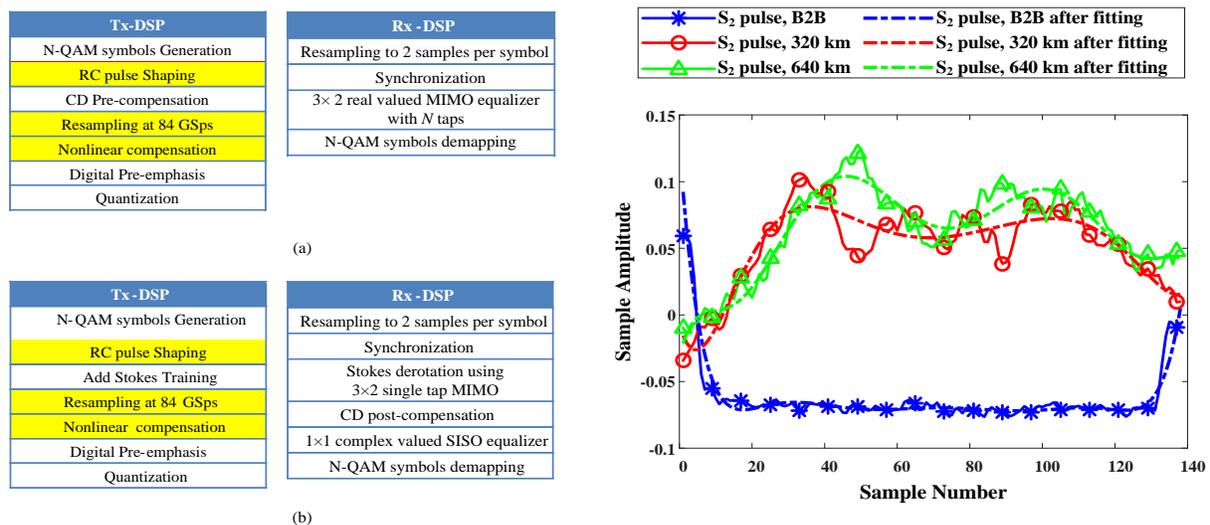


Fig. 5.5 Transmitter and receiver DSP stacks for two CD compensation methods: (a) CD pre-compensation, and (b) CD post-compensation. The inset on the right shows one long Stokes training pulse in B2B and after 320 and 640 km in case of a 84 Gbaud signal.

In the 64 Gbaud case, the Tx-DSP is the same as the Tx-DSP used in subsection 5.4.1 where the RC pulse shaping filter has roll-off factor of 0.03. Also, the resampling process is done at new sampling rate of 84 GSps after CD pre-compensation block. Digital pre-emphasis is applied after non-linear compensation block where the samples are pre-distorted using an FIR filter optimized experimentally to enhance the system performance by properly balancing the amount of intersymbol interference (ISI) pre-compensated at the transmitter and the residual ISI left to be post-compensated at the receiver [59]. Finally, the pre-distorted samples are quantized to 256

levels and uploaded to the memory of the 8-bit DAC. In the 84 Gbaud case, the yellow shaded blocks in the Tx-DSP stack of Fig. 5.5(a) are not used since neither pulse shaping nor resampling is needed. Also, the nonlinear compensation block is not used in the 84 Gbaud case because the output swing (V_{pp}) from the RF amplifiers is less than 2 V which almost lies in the linear region of the MZM.

When CD post-compensation is adopted, the Tx-DSP, shown in Fig. 5.5(b), differs from the one shown in Fig. 5.5(a) when CD pre-compensation is used in one block: the CD pre-compensation block is removed and replaced by the block in which Stokes training symbols are added. These training symbols are used by the Rx-DSP to perform the Stokes polarization de-rotation before applying the CD post-compensation, which is essential as pointed out in section 5.2.2. In the 64 Gbaud case, the training rate is one Gbaud, i.e., the duration of a training symbol is 64 times the data symbol duration. The structure of the training sequence is such that (S_1, S_2, S_3) are sent in the following sequence (1,0,0), (0,1,0), (0,-1,0), (0,0,1), and (0,0,-1). This sequence is repeated 40 times, and the sign inversion used in the S_2 and S_3 parameters is due to the AC coupled nature of the DAC. In the 84 Gbaud case, the same training sequence explained above is used while the training rate is changed to 1.3125 Gbaud, when the transmitted distance is less than 960 km. The training rate is lowered to 0.6525 Gbaud, when the transmitted distance is 960 km. This reduction of the training rate is to ensure the training sequence is CD tolerant to this long distance.

The offline receiver DSP (Rx-DSP) stacks in both Fig. 5.5(a) and Fig. 5.5(b) start by resampling the captured patterns from 80/160 GSps to twice the data symbol rate. The next block in the stack of Fig. 5.5(a) is a synchronization process in order to locate the training symbols (TS) required to start the tap adaptation of the subsequent 3×2 MIMO equalizer using training symbol based least mean squares algorithm (TS-LMS). It is noteworthy that these training symbols are transmitted at the same symbol rate of the data payload since CD pre-compensation is employed. After the TS period, decision directed LMS (DD-LMS) replaces TS-LMS in steady-state operation. The 3×2 MIMO equalizer is a bank of real valued adaptive filters, each having N taps, that performs two roles in one stage: it tracks the polarization rotation to retrieve the transmitted Stokes parameters S_2^t and S_3^t and equalizes for any residual inter-symbol interference (ISI). On the other hand, if CD post-compensation is used, a synchronization block is also required in the stack of Fig. 5.5(b) in order to locate the special slow Stokes training

symbols that are used to estimate the polarization de-rotation matrix. The right inset plot in Fig. 5.5 shows the training pulse experimentally captured in case of 84 Gbaud at twice the symbol rate, i.e., the pulse is represented in 128 samples after transmission of 320 km (in red) and 640 km (in green). The B2B pulse is overlaid inverted in blue to clarify the effect of the CD on the pulse. It can be seen that the middle samples of the Stokes training symbol pulse are not affected by CD and can be used to estimate the Stokes de-rotation matrix. The number of these CD-tolerant samples decreases with the increase of the transmission distance to the point that the training symbol rate may need to be reduced for better estimation. The next block applies this estimated de-rotation matrix, which can be viewed as a 3×2 MIMO of single tap filters, to perform polarization de-rotation. Only at this point of the Rx-DSP is where CD can be post-compensated. Subsequently, residual ISI can be mitigated by via a complex valued 1×1 SISO equalizer that provides an estimate of $S_2^t + jS_3^t$. Finally, the received symbols are de-mapped into bits for BER counting.

In summary, there are three advantages of the CD pre-compensation over the CD post-compensation: (1) it omits the need of slow Stokes training symbols per frame, (2) the performance of the polarization tracking is independent from the transmitted distance, and (3) it allows the merging of two DSP stages in one stage at the Rx-DSP. On the other hand it CD pre-compensation suffers from the following two disadvantages: (1) it requires prior knowledge of the transmission distance, and (2) it increases the PAPR of the driving electrical signals at the transmitter due to pre-distortion which may degrade the resulting signal quality.

5.5 Results and discussion

5.5.1 First phase of 2D-SV-DD system demonstration

Fig. 5.6 shows the BER versus transmitted distance for a 56 Gbaud signal at optimum CSCR values for QPSK, 8QAM and 16QAM formats. The optimum launch power, which is experimentally found, is 5 dBm for QPSK and 6 dBm for 8QAM and 16QAM. At the optimum launch power and CSCR values, we achieve the following three key transmission results with BER below the HD-FEC threshold using the 2D-SV-DD: i) 320 km with 224 Gb/s 16QAM, ii) over 1000 km with 168 Gb/s 8QAM, and iii) 2880 km with 112 Gb/s QPSK.

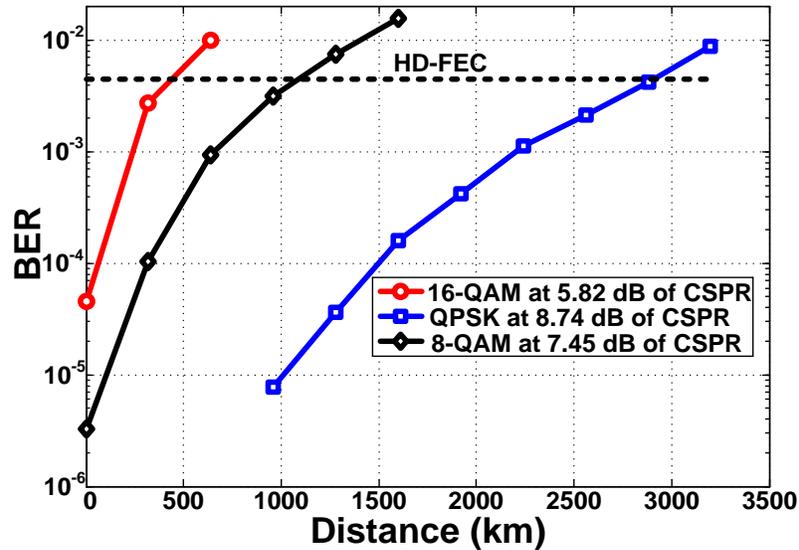


Fig. 5.6 BER versus transmission distance at the optimum CSPR and the optimum launch power for QPSK, 16QAM, and 8QAM modulation formats at 56 Gbaud using 2D-SV-DD.

Next, Fig. 5.7 shows the CSPR optimization results for all three modulation formats. More specifically, it depicts the BER versus CSPR at the maximum reach where BER hits the HD-FEC threshold for a given format. The CSPR measurement is done by measuring the output power from the modulator before and after loading the waveforms into the DAC for each format with corresponding CD pre-compensation. The measured output power from the modulator gives either: a) the carrier power going through the Y polarization of the DP-IQ modulator when the modulating RF waveforms of the X-polarization are zero, b) the total power for the signal and carrier when the modulating RF waveforms are loaded into the DAC. In this SV-DD configuration, the carrier on Y polarization is the phase diversity reference for the modulated signal on the X polarization, which is equivalent to the role of the local oscillator (LO) in a coherent system. Thus, CSPR is an important parameter to investigate, similar to LO-to-Signal-Power-Ratio (LSPR) at the receiver in coherent transmission [54, 141]. From Fig. 5.7, as the QAM order increases, the optimum CSPR value decreases, giving more power on the “signal” polarization. This is because the total launch power from the booster for the modulated signal and the carrier is constant and increasing the QAM order, i.e., having a denser constellation, requires inherently larger OSNR for detecting the signal itself. This means that the required modulated signal launch power increases which results in a lower optimum CSPR value.

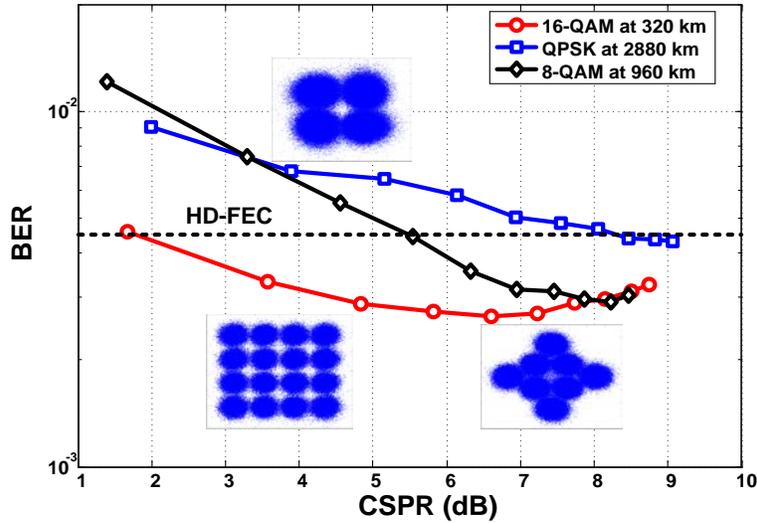


Fig. 5.7 BER versus CSRP of various format at fixed transmitted distance with constellations.

Finally, Table 5-1 compares system requirements (hardware and DSP functions) and achieved performance of this current work against several recent demonstrations of SV-DD systems targeting short reach and metro applications [78, 79, 83] as well as a comparison with a recent InP IQM enabled 400G coherent system demonstration for metro applications [59] for benchmarking purposes. Coherent systems enable modulation of four available dimensions for modulation, namely the amplitude and phase of the electric field on two orthogonal polarizations. However, it has the highest system complexity in terms of required DSP and transceiver components as indicated in Table 5-1. On the other hand, though sharing the same capacity as SV-DD system, single polarization (SP) coherent transmission requires either a polarization stabilizer to align the randomly varying SOP in the fiber to that of the LO, or a polarization diversity configuration consisting of a 2×8 dual polarization hybrid and four BPDs [73].

In [83], a dual polarization intensity modulation (DP-IM) system is demonstrated for short reach applications enabling two-dimensional modulation (2D modulation) which doubles the number of bits per symbol compared to conventional IM/DD with one dimensional modulation. This approach is suitable for short reach applications where reach is less than 20 km and operating in the O-band is essential to avoid the effects of CD. In [79], the authors used an OFDM modulation technique in SV-DD system where a single polarization IQ modulator (SP-IQM) is used on one polarization and the carrier is transmitted on the orthogonal polarization. Thus, CD post-compensation can be done in such a system after the polarization rotation is

inverted following the method described in [81] which extends the reach of SV-DD systems. In our work, we use the same configuration in [79] (complex modulation on one polarization and carrier transmission on the other polarization), but instead we use single carrier modulation with CD pre-compensation at the transmitter replacing CD post-compensation. Our 2D-SV-DD scheme shows a significant reduction in terms of system complexity (hardware and DSP functions) compared to coherent systems while maintaining an almost similar performance of a SP coherent system.

Table 5-1 Schemes for SV-DD compared to 400G coherent system demonstration (SP-IQM: single polarization IQ modulator, IFFT: inverse fast Fourier transform)

	DP coherent system	SV-DD system (DP-IM)	SV-DD system SP-IQ modulation	Our work
Tx main components	<ul style="list-style-type: none"> • DP-IQM (6 MZMs) • 4 DAC channels • 4 RF amplifiers 	<ul style="list-style-type: none"> • 2 MZMs • 2 DAC channels • 2 RF amplifiers 	<ul style="list-style-type: none"> • SP-IQM (3 MZMs) • 2 DAC channels • 2 RF amplifiers 	<ul style="list-style-type: none"> • SP-IQM (3 MZMs) • 2 DAC channels • 2 RF amplifiers
Rx main components	<ul style="list-style-type: none"> • Dual Pol. Hybrid • 4 BPDs • 4 ADCs • LO 	<ul style="list-style-type: none"> • Single Pol. Hybrid • 2 SE-PDs and 2 BPDs • 4 ADCs 	<ul style="list-style-type: none"> • Single Pol. Hybrid • 2 SE-PDs and 2 BPDs • 4 ADCs 	<ul style="list-style-type: none"> • Single Pol. Hybrid • 2 SE-PDs and 2 BPDs • 4 ADCs (3 ADCs if balanced detection used to construct of S_I parameter)
Main use of DSP at transmitter and receiver	<ul style="list-style-type: none"> • Digital Pre-emphasis • 2x2 complex MIMO • CD post-comp • Freq. Offset Removal • Phase noise mitigation 	<ul style="list-style-type: none"> • Digital Pre-emphasis • 4x2 real MIMO 	<ul style="list-style-type: none"> • OFDM synch • Stokes training (4x2 real MIMO) • IFFT / FFT • Channel equal. • CD post-comp 	<ul style="list-style-type: none"> • Digital Pre-emphasis • CD pre-comp • 3x2 real MIMO
Modulation format	DP-16QAM	DP-PAM4	OFDM	16QAM
Modulated dimensions	4	2	2	2
Throughput-times-distance (Gb/s.km)	143,360	2,240	48,000	71,680
Bit rate (Gb/s)	448	224	100	224
Reach (km)	320	10	480	320
BER at specified reach	1.5×10^{-3}	3.8×10^{-3}	$\sim 2 \times 10^{-2}$	2.8×10^{-3}
Reference	[59]	[83]	[79]	[58]

5.5.2 Second phase for CD impact on 2D-SV-DD demonstration

Fig. 5.8 shows the BER versus the transmitted distance to compare the CD pre-compensation and the CD post-compensation at 84 Gbaud and 64 Gbaud 16QAM cases. In case of the 84 Gbaud 16QAM, the CD post-compensation has a better performance than the CD pre-compensation where the BER is $\sim 1.67 \times 10^{-2}$ and $\sim 2 \times 10^{-2}$ in case of CD post- and pre-compensation, respectively. In case of the 64 Gbaud, the situation is inverted where the CD pre-compensation performs better than the CD post-compensation where the BER at 960 km is $\sim 1.61 \times 10^{-2}$ and $\sim 9.8 \times 10^{-3}$ in case of CD post- and pre-compensation, respectively. This can be attributed to two reasons; first, in case of 64 Gbaud, we use a small roll-off factor (0.01) to confine the signal bandwidth in the 33 GHz bandwidth of the RTO. This means that the 64 Gbaud signal bandwidth is also less than Nyquist frequency at the transmitter which is half the DAC sampling rate (42 GHz). Thus, we can fully pre-compensate for the CD at the Tx-DSP over the entire bandwidth of the 64 Gbaud signal in the frequency domain. However, in case of 84 Gbaud signal, the symbol rate equals to the DAC sampling rate and pulse shaping is not possible to have a band limited 84 Gbaud signal. Consequently, there will be some spectral content above the 42 GHz for which CD pre-compensation is not possible, i.e., CD pre-compensation operates properly only up to the Nyquist frequency and any information content above this will be impacted by CD. On the other hand, the higher sampling rate of the RTO allows for higher oversampling factor enabling the CD post-compensation to compensate CD over a larger signaling bandwidth and hence outperform CD pre-compensation. Second, the effect of the pre-distortion on the output electrical signal quality of the DAC in case of CD pre-compensation while operating at the DAC sampling rate increases the PAPR of the output signal from the DAC increasing the DAC requirements in terms of ENOB and quantization noise to fully represent the pre-distorted signal.

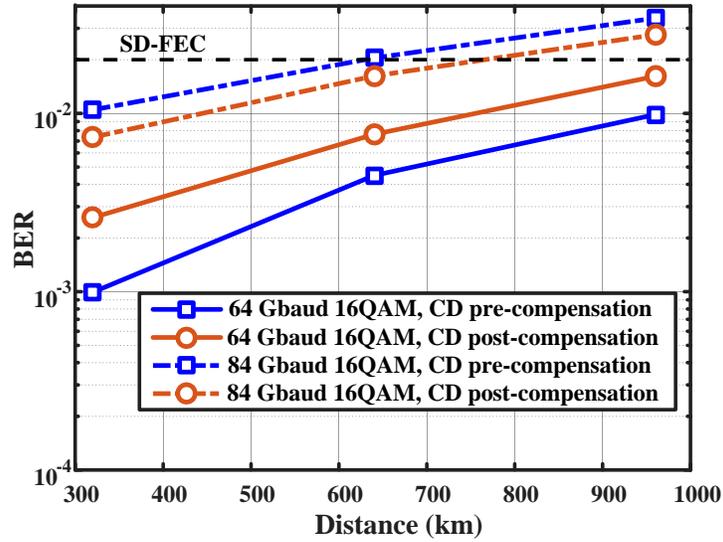


Fig. 5.8 BER versus distance (CD pre-compensation and post-compensation comparison at 84 Gbaud and 64 Gbaud 16QAM cases).

Fig. 5.9 shows the BER versus CSRR optimization for the 64 Gbaud 16QAM and 84 Gbaud 16QAM in case of the CD pre-compensation at 960 km and 640 km, respectively. The optimum total launched power is 6.5 dBm and 5.5 dBm and the optimum CSRR is 7.2 dB and 8 dB, in case of 64 Gbaud 16QAM and 84 Gbaud 16 QAM, respectively.

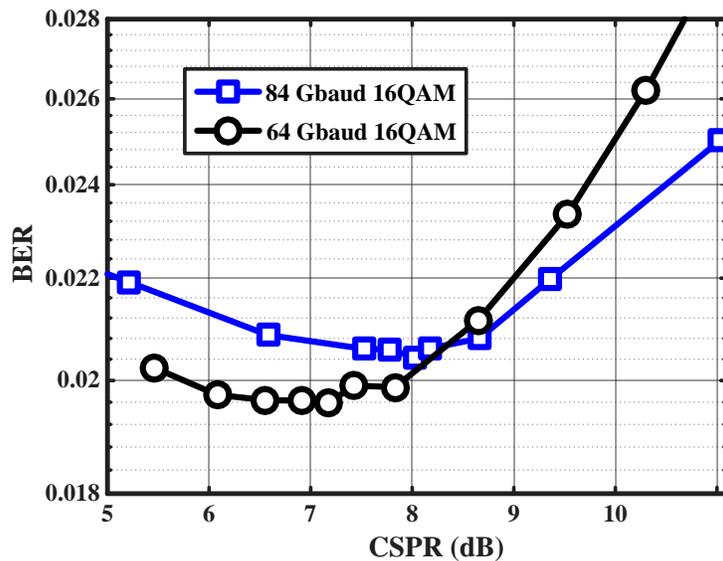


Fig. 5.9 BER versus CSRR for 84 Gbaud 16QAM and 64 Gbaud 16QAM using CD pre-compensation.

As mentioned in the DSP stack subsection 5.4.2, one more advantage of the CD pre-compensation is the extended reach where in case of the CD post-compensation the Stokes training symbols has to be slower as the reach increases. Fig. 5.10 shows the BER versus distance to evaluate the 2D-SV-DD system performance using CD pre-compensation at symbol rates of 84 Gbaud and 64 Gbaud using different modulation formats (16QAM, 8QAM and QPSK) at optimized launched power and CSNR. The maximum reach is 640 km, 1280 km, and 3520 km at SD-FEC threshold in case 84 Gbaud 16QAM, 8QAM, and QPSK, respectively. This extended reach especially for 8QAM and QPSK cannot be achieved by CD post-compensation using the Stokes training at 1 Gbaud and slower training will be affected by the AC coupling of the transmitter electronics. This will be more difficult In the case of 64 Gbaud where the maximum reach is 1280, 2240, and 4800 km at SD-FEC threshold using the modulation formats 16QAM, 8QAM, and QPSK, respectively.

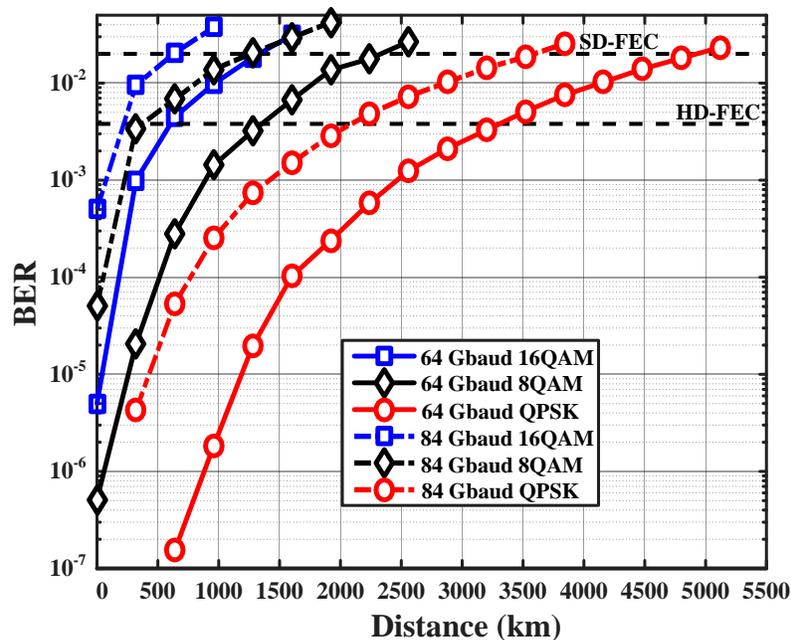


Fig. 5.10 BER versus distance in case of CD pre-compensation for signals with symbol rates of 64 Gbaud and 84 Gbaud at different modulation formats

5.6 Conclusion

We demonstrated 100G and 200G Stokes-Vector transmission using an InP DP-IQM and an SVR DD on a single carrier using different modulation formats, namely QPSK, 8QAM, and 16QAM. For 56 Gbaud 16QAM signal, the total throughput-times-distance product is

71,680 Gb/s.km at an optimized CSPR value of 6 dB. For 56 Gbaud QPSK signal, the total throughput-times-distance product is 322,560 Gb/s.km at an optimized CSPR value of 8 dB. These results show the viability of using SVR DD and small form factor integrated InP IQM to meet 100G and 200G transmission system requirements for inter-datacenter applications.

In addition, we demonstrated the 2D-SV-DD system at symbol rates of 84 Gbaud and 64 Gbaud with different modulation formats using CD pre-compensation. We showed analytically the importance of performing the Stokes de-rotation before applying the CD post-compensation on the S_2 and S_3 parameters. The CD pre-compensation omits the need of the Stokes training symbols which reduces the overall data rate, allows for merging the Stokes de-rotation with the ISI mitigation in single stage 3×2 real valued MIMO at Rx-DSP, and is more suitable for extended reaches with respect to the CD post-compensation method. On the other hand, CD pre-compensation affects the transmitted signal from the DAC, where this pre-distortion increases the PAPR of the transmitted samples which increases the quantization noise and requires high ENOB to sufficiently represent the CD pre-compensated samples. This effect can be noticed at high symbol rate, i.e., at one sample per symbol, and high order of modulation format as in the case of 84 Gbaud 16 QAM. Despite this effect, the CD pre-compensation outperforms the CD post-compensation in the 64 Gbaud case. Also, the CD pre-compensation method allows for extended reach where we are able to demonstrate the highest reported throughput-times-distance products of 591,360 Gb/s.km and 614,400 Gb/s.km in Stokes transmission using QPSK modulation format and symbol rates of 84 Gbaud and 64 Gbaud, respectively.

Chapter 6

Conclusion and Future Work

6.1 Summary of presented work

In **Chapter 2**, we have proposed a new contention resolution strategy in OBS networks for intra-datacenter networking, where we have used the electronic assembly buffer to provide two tasks; (i) constructing the bursts from the aggregated traffic of packets having the same destination, and (ii) resolving any potential contention by keeping the bursts in the electronic assembly buffer until the contending burst is served. This omits the need of virtual output queues (VOQs) at the core switch. In addition, we have proposed an AWGR-based switching mechanism that achieves with the aid of the centralized scheduling and locality of pods inside a datacenter totally passive core switching where fast TLs are used at the transmitter for burst transmission and for wavelength routing at the same time. Finally, we have investigated the impact of the self-similar characteristics of the network traffic on the intra-datacenter network performance. OPS provides lower latency at low degree of self-similarity and/or at low traffic load because of the additional latency due to assembly needed in OBS. OBS shows only slight improvement in terms of latency over OPS at high traffic loads and high degree of self-similarity. Thus, the main advantage of OBS is the reduction in the number of requests received by the scheduler due to lower number of bursts compared to the number of packets at the same offered load.

In **Chapter 3**, we have proposed a self-homodyne system for next generation DCIs in intra-datacenter networking. The proposed system takes advantage of the target application to send the tone and the modulated signal on two different fiber lanes of a duplex fiber where the tone can be used for coherent reception. The proposed system have several advantages that can be outlined as follows:

- a. The reduction of the Rx-DSP relative to a conventional coherent system.
- b. The capacity offered by the proposed system per wavelength is higher than that of the IM/DD system.
- c. Given the same overall target capacity required, the physical complexity of the proposed system approaches the corresponding IM/DD system.

- d. The ability of operation in C-band and O-band without noticeable additional complexity due to CD.
- e. The proposed system is colorless, i.e., it is immune to laser frequency drifts. This enables the deployment of uncooled lasers.

In addition, we have demonstrated the proposed self-homodyne system in a proof-of-concept experiment using an integrated InP DP-IQM. We report, to the best of our knowledge, record breaking transmissions of 530 Gb/s, 448 Gb/s, and 320 Gb/s signals on a single wavelength over 500 m, 2 km, and 10 km of SMF fibers, respectively, below KP4 threshold using integrated InP modulator. These bit rates reported align well with the next generation target capacities of intra-datacenter interconnects.

In **Chapter 4**, we have investigated the viability of using the InP DP-IQM for 400G system inter-datacenter reaches. We demonstrated experimentally 400G transmission using 65.7 GSps DAC and an InP DP-IQM on a single carrier including the required overhead at different modulation formats and symbol rates, namely 16QAM at 56 Gbaud and 64QAM at both 38 and 43 Gbaud. Transmission of the 56 Gbaud 16QAM signal needs lower OSNR and can propagate over longer distances (more than 500 km) compared to 64QAM signals at lower symbol rates. Moreover, we pushed the limits of the device under test and demonstrated 770 Gb/s transmission using 84 GSps DAC on a single carrier at a symbol rate of 77 Gbaud and PDM-32QAM modulation format over 320 km below SD-FEC threshold. In addition, transmissions are demonstrated at a symbol rate of 84 Gbaud, where bit rates of 504 Gb/s and 672 Gb/s are achieved using 8QAM and 16QAM modulation formats over 960 km and 1920 km, respectively below SD-FEC threshold. Results in both demonstrations show that the OSNR implementation penalty is highly dependent on the order of the modulation format due to its sensitivity to the transmitter noise.

In **Chapter 5**, we have investigated the 2D-SV-DD system as a potential alternative to conventional coherent system for inter-datacenter interconnects. We experimentally demonstrated 100G and 200G Stokes-Vector transmission using an InP DP-IQM and an SVR DD on a single carrier. For 56 Gbaud 16QAM signal, the total throughput-times-distance product is 71,680 Gb/s.km at an optimized CSPR value of 6 dB. For 56 Gbaud QPSK signal, the total throughput-times-distance product is 322,560 Gb/s.km at an optimized CSPR value of 8 dB. These results show the viability of using SV-DD and small form factor integrated InP IQM to

meet 100G and 200G transmission system requirements for inter-datacenter applications. In addition, we have demonstrated the 2D-SV-DD system at symbol rates of 84 Gbaud and 64 Gbaud with different modulation formats using CD pre-compensation. Using CD pre-compensation allows for extended reach where we are able to demonstrate the highest reported throughput-times-distance products of 591,360 Gb/s.km and 614,400 Gb/s.km in Stokes transmission using QPSK modulation format and symbol rates of 84 Gbaud and 64 Gbaud, respectively. In addition, we have shown analytically the importance of performing the Stokes de-rotation before applying the CD post-compensation on the S_2 and S_3 parameters. The CD pre-compensation omits the need of the Stokes training symbols which reduces the overall data rate, allows for merging the Stokes de-rotation with the ISI mitigation in single stage 3×2 real valued MIMO at Rx-DSP, and is more suitable for extended reaches with respect to the CD post-compensation method. Finally, we deduced from the CD post- and pre-compensation comparison results that the CD pre-compensation affects the transmitted signal quality from the DAC, where this pre-distortion increases the PAPR of the transmitted samples which increases the quantization noise and requires high ENOB to sufficiently represent the CD pre-compensated samples. This effect can be highly noticed at large symbol rate, i.e., at one sample per symbol, and high order of modulation formats like in the case of 84 Gbaud 16 QAM.

6.2 Future research avenues

In this section, we explore the potential research directions that can be pursued in each of the topics covered by the thesis. Also, we highlight the possible applications in the future for next generation networks in datacenters.

6.2.1 Potential research extension for presented work

Optical switching for intra-datacenter networking

- Extend the work to investigate the physical layer performance of the proposed AWGR-based system. This is important because the AWGR will not receive and retransmit the data while electronic switching can be considered as a regenerating station in addition to the switching task. Also, if neighboring channels are used at the same time, the probability of the cross talk will increase. In the proposed network architecture, only one

case will be available when the last channel in an FSR and the first channel of the next FSR are used at the AWGR input/output.

- Propose an intelligent adaptive solution to allow for hybrid OPS/OBS implementation. The main objective is to realize this adaptive solution without traffic pattern analysis. As concluded in section 2.7, the OPS has a better performance than OBS at low traffic loads. The OBS slightly improves the network. The main advantage of the OBS is the reduced number of requests to be handled at the scheduler relative to OPS especially at high traffic loads. Thus, the potential of having hybrid solution to realize optical switching motivates the research in that direction.
- Investigate a solution to combine the proposed self-homodyne system in Chapter 3 with the proposed AWGR-based switching mechanism in Chapter 2. This needs investigation in terms of network architecture to do the switching not only for the data available on the modulated signal, but also for the transmitted tone used for coherent detection. Also, the system performance of the self-homodyne system needs to be evaluated along double the distances demonstrated in Chapter 3 because the AWGR switch is a passive optical switch.

Self-homodyne system for Intra-datacenter networking

- After the proof of concept using emulation, the second logical step is to investigate a real system with full-duplex fiber deployment. This will enable the flexible use of wavelengths at both transceivers, i.e., we can study when both transceivers use the same wavelength for transmission or different wavelengths. This study should include the fiber nonlinear effects on the system performance in two cases when the counter-propagating tone and signal have the same wavelength and when they operate at different wavelengths at different channel spacing.
- In the emulation described in Chapter 3, we used a linear driver with limited bandwidth. The performance of this system can be further enhanced using amplifiers with higher bandwidth. This will enable the increase of the used symbol rates while keeping the use of the KP4 threshold as a target BER after transmission.
- Investigate an integrated solution for the polarization stabilizer required at the receiver. The main target is to design a polarization stabilizer that does not require any feedback

circuits to control the polarization stabilizer, i.e., a passive polarization stabilizer is desirable with low delay and compact in size.

Optical solutions for inter-datacenter networking

- Since the InP modulator is characterized by low values of V_π , we can investigate the performance of a driverless transmitter, i.e., removing the RF amplifiers prior to the DP-IQM, which would improve the RF signal quality and reduce the power consumption.
- The digital pre-emphasis we used up till now is dedicated to each channel. We can investigate the viability of allowing for MIMO digital pre-emphasis per polarization (two 2×2 MIMO blocks).
- Investigate the viability of 3D-SV-DD transmission system over extended distances with respect to available solutions in literature [78]. All of the available 3D-SV-DD systems are proposed for short reach applications. We can use the DP-IQM in 3D-SV-DD system enabling the CD pre-compensation on both polarizations. This may increase the capacity per wavelength relative to 2D-SV-DD and keep the reduced complexity of the receiver relative to coherent system.

6.2.2 Roadmap for future implementation

Finally, we would like to highlight what we are expecting regarding the implementation of the work presented in this thesis.

The work presented in Chapter 2 can be considered as a milestone that utilizes the collaboration of the system entities for the implementation of all-optical switching in the core layer in datacenters. It is not expected that all-optical switching can be used in the access layer level within the next ten years. However, there is a promising avenue to apply all-optical switching in the core layer switching. Several aspects need to be considered replacing electronic switching in terms of port count, latency, and throughput. The electronic switching is more advanced than optical switching in several aspects including the multi-stage switching, the adaptation with traffic conditions with multiple routing options. When these important aspects are handled, the implementation of optical switching will be considered. I believe that this can only happen when the industry starts to invest towards this research with collective

specializations. This investment will be a must after 2020, when the required data rates reach beyond 1.6 Tb/s according to [36].

The work presented in Chapter 3 is highly promising work especially with the continuous increase in the data rate demands. I believe within the next few years the coherent system will migrate from transport layer applications to the short reach applications in the 2 km and 10 km reaches as it has dominated the long-haul applications. The commodity of integrated coherent transceivers is related to the demand of its implementation in short reach applications. If it is applied in the short reach applications, the mass production of the coherent transceivers with reduced complexity will make this system realizable in the next five years. Research needs to be done to further reduce the system complexity like realizing a coherent receiver with a front-end that omits the need of the polarization demultiplexing DSP block which can be considered as the only remaining DSP block that is different from the DSP stack used in the IM/DD system.

Regarding the inter-datacenter reaches, the coherent system presented in Chapter 4 is already deployed in those reaches. Thus, our work in that chapter is already relevant in terms of enhancing the capacity specific to the inter-datacenter reaches with the deployment of integrated solutions with high symbol rates and advanced modulation formats. The work presented in Chapter 5 has a potential application in distances with moderate reaches like 80 km and few hundreds of kilometers [81]. This is because it has almost the same capacity of a single polarization coherent system with reduced system complexity. However, I believe that the full dual-polarization coherent system will be dominant in most of the inter-datacenter reaches due to double the capacity that is provided over the 2D-SV-DD system.

Bibliography

- [1] Cisco Global Cloud Index: Forecast and Methodology, 2015–2020 , White-Paper-c11-738085. Available: <http://www.cisco.com/c/dam/en/us/solutions/collateral/service-provider/global-cloud-index-gci/white-paper-c11-738085.pdf>
- [2] A. Erik *et al.*, “Roadmap of optical communications,” *Journal of Optics*, vol. 18, no. 6, pp. 063002-063041, May 2016.
- [3] D. Liang and J. E. Bowers, “Photonic integration: Si or InP substrates?,” *Electronics Letters*, vol. 45, no. 12, pp. 578-581, June 2009.
- [4] Y. Akulova, “Advances in integrated widely tunable coherent transmitters,” in *Proc. OFC 2016*, Anaheim, CA, 20-24 Mar. 2016, p. W4H.1.
- [5] M. K. Smit, “Generic InP-based photonic integration technology,” in *Proc. OECC 2012*, Busan, 2-6 July 2012, pp. 289-290.
- [6] M. Smit *et al.*, “An introduction to InP-based generic integration technology,” *Semiconductor Science and Technology*, vol. 29, no. 8, pp. 083001-083041, June 2014.
- [7] G. P. Agrawal, *Lightwave technology: Telecommunication systems*. Wiley, 2005.
- [8] J. P. Jue and V. M. Vokkarane, *Optical burst switched networks*. Springer Science & Business Media, 2006.
- [9] X. Yu, Y. Chen, and C. Qiao, “Study of traffic statistics of assembled burst traffic in optical burst-switched networks,” in *in Proc. OptiComm 2002*, Boston, MA, 5 July 2002, pp. 149-159.
- [10] X. Yu, J. Li, X. Cao, Y. Chen, and C. Qiao, “Traffic statistics and performance evaluation in optical burst switched networks,” *IEEE/OSA Journal of lightwave technology*, vol. 22, no. 12, pp. 2722-2738, Dec. 2004.
- [11] T. Battestilli and H. Perros, “An introduction to optical burst switching,” *IEEE communications magazine*, vol. 41, no. 8, pp. S10-S15, Aug. 2003.
- [12] Y. Chen, C. Qiao, and X. Yu, “Optical burst switching: a new area in optical networking research,” *IEEE Network*, vol. 18, no. 3, pp. 16-23, June 2004.
- [13] I. Baldine, G. N. Rouskas, H. G. Perros, and D. Stevenson, “JumpStart: A just-in-time signaling architecture for WDM burst-switched networks,” *IEEE communications magazine*, vol. 40, no. 2, pp. 82-89, Feb. 2002.

- [14] M. Yoo and C. Qiao, "Just-enough-time (JET): A high speed protocol for bursty traffic in optical networks," in *Proc. IEEE/LEOS Summer Topical Meeting*, Montreal, QC, Canada, 11-13 Aug. 1997, pp. 26-27.
- [15] S. D. Lucente, N. Calabretta, J. A. C. Resing, and H. J. S. Dorren, "Scaling low-latency optical packet switches to a thousand ports," *IEEE/OSA Journal of Optical Communications and Networking*, vol. 4, no. 9, pp. A17-A28, Sept. 2012.
- [16] R. Hemenway, R. Grzybowski, C. Minkenberg, and R. Luijten, "Optical-packet-switched interconnect for supercomputer applications [Invited]," *Journal of Optical Networking*, vol. 3, no. 12, pp. 900-913, Jan. 2004.
- [17] S. B. Yoo, "Optical packet and burst switching technologies for the future photonic internet," *IEEE/OSA Journal of Lightwave Technology*, vol. 24, no. 12, pp. 4468-4492, Jan. 2006.
- [18] N. Farrington *et al.*, "Helios: a hybrid electrical/optical switch architecture for modular data centers," *ACM SIGCOMM Computer Communication Review*, vol. 40, no. 4, pp. 339-350, Oct. 2010.
- [19] Y. Ueda *et al.*, "4×4 InAlGaAs/InAlAs optical switch fabric by cascading Mach-Zehnder interferometer-type optical switches with low-power and low-polarization-dependent operation," *IEEE Photonics Technology Letters*, vol. 24, no. 9, pp. 757-759, May 2012.
- [20] B. G. Lee *et al.*, "Monolithic silicon integration of scaled photonic switch fabrics, CMOS logic, and device driver circuits," *IEEE/OSA Journal of Lightwave Technology*, vol. 32, no. 4, pp. 743-751, Feb. 2014.
- [21] A. V. Rylyakov *et al.*, "Silicon photonic switches hybrid-integrated with CMOS drivers," *IEEE Journal of Solid-State Circuits*, vol. 47, no. 1, pp. 345-354, Jan. 2012.
- [22] Y. Xiaohui, P. Mejia, Y. Yawei, R. Proietti, S. J. B. Yoo, and V. Akella, "DOS - A scalable optical switch for datacenters," in *Proc. ANCS 2010*, La Jolla, CA, 25-26 Oct. 2010, pp. 1-12.
- [23] Q. Xu, H. Rastegarfar, Y. B. M'Sallem, A. Leon-Garcia, S. LaRochelle, and L. A. Rusch, "Analysis of large-scale multi-stage all-optical packet switching routers," *IEEE/OSA Journal of Optical Communications and Networking*, vol. 4, no. 5, pp. 412-425, June 2012.

- [24] R. Proietti *et al.*, “All-optical NACK for fast packet retransmission in AWGR-based optical switches,” in *Proc. ECOC 2011*, Geneva, 18-22 Sept. 2011, p. We.10.P1.50.
- [25] X. Ye, R. Proietti, Y. Yin, S. Yoo, and V. Akella, “Buffering and flow control in optical switches for high performance computing,” *IEEE/OSA Journal of Optical Communications and Networking*, vol. 3, no. 8, pp. A59-A72, Aug. 2011.
- [26] Y. Yin, R. Proietti, X. Ye, C. J. Nitta, V. Akella, and S. J. B. Yoo, “LIONS: An AWGR-based low-latency optical switch for high-performance computing and data centers,” *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 19, no. 2, pp. 3600409-3600409, Apr. 2013.
- [27] S. Qiu *et al.*, “A cost-effective scheme of high-radix optical burst switch based on fast tunable lasers and cyclic AWG,” in *Proc. OFC 2012*, Los Angeles, CA, 4-8 Mar. 2012, p. OTh1G.4.
- [28] Q. Huang, Y.-K. Yeo, and L. Zhou, “A single-stage optical load-balanced switch for data centers,” *OSA Optics express*, vol. 20, no. 22, pp. 25014-25021, Oct. 2012.
- [29] C. Guo *et al.*, “BCube: a high performance, server-centric network architecture for modular data centers,” *ACM SIGCOMM Computer Communication Review*, vol. 39, no. 4, pp. 63-74, Oct. 2009.
- [30] G. Wang *et al.*, “c-Through: Part-time optics in data centers,” *ACM SIGCOMM Computer Communication Review*, vol. 40, no. 4, pp. 327-338, Oct. 2010.
- [31] C. Guo, H. Wu, K. Tan, L. Shi, Y. Zhang, and S. Lu, “Dcell: a scalable and fault-tolerant network structure for data centers,” *ACM SIGCOMM Computer Communication Review*, vol. 38, no. 4, pp. 75-86, Oct. 2008.
- [32] L. Peng, C.-H. Youn, W. Tang, and C. Qiao, “A novel approach to optical switching for intradatacenter networking,” *IEEE/OSA Journal of Lightwave Technology*, vol. 30, no. 2, pp. 252-266, Jan. 2012.
- [33] M. Imran, M. Collier, P. Landais, and K. Katrinis, “Software-defined optical burst switching for HPC and cloud computing data centers,” *IEEE/OSA Journal of Optical Communications and Networking*, vol. 8, no. 8, pp. 610-620, Aug. 2016.
- [34] *IEEE Standard for Ethernet Amendment 2: Physical Layer Specifications and Management Parameters for 100 Gb/s Operation Over Backplanes and Copper Cables*, 2014.

- [35] *IEEE P802.3bs 400GbE Task Force Public Area*. Available: http://www.ieee802.org/3/bs/public/15_03/index.shtml
- [36] *Ethernet Alliance Roadmap*. Available: <http://www.ethernetalliance.org/roadmap/>
- [37] M. Poulin *et al.*, "107 Gb/s PAM-4 transmission over 10 km using a SiP series push-pull modulator at 1310 nm," in *Proc. ECOC 2014*, Cannes, France, 21-25 Sept. 2014, p. Mo.4.5.3.
- [38] K. Zhong *et al.*, "140 Gbit/s 20km transmission of PAM-4 signal at 1.3 μm for short reach communications," *IEEE Photon. Technol. Lett.*, vol. 27, no. 16, pp. 1757-1760, Aug. 2015.
- [39] M. Chagnon, M. Morsy-Osman, M. Poulin, C. Paquet, S. Lessard, and D. V. Plant, "Experimental parametric study of a silicon photonic modulator enabled 112-Gb/s PAM transmission system with a DAC and ADC," *IEEE/OSA Journal of Lightwave Technology*, vol. 33, no. 7, pp. 1380-1387, Apr. 2015.
- [40] K. Zhong *et al.*, "Experimental study of PAM-4, CAP-16, and DMT for 100 Gb/s short reach optical transmission systems," *OSA Optics Express*, vol. 23, no. 2, pp. 1176-89, Jan 2015.
- [41] X. Chen *et al.*, "All-electronic 100-GHz bandwidth digital-to-analog converter generating PAM signals up to 190 GBaud," *IEEE/OSA Journal of Lightwave Technology*, vol. PP, no. 99, pp. 1-1, 2016.
- [42] H. Mardoyan *et al.*, "84-, 100-, and 107-GBd PAM-4 intensity-modulation direct-detection transceiver for datacenter interconnects," *IEEE/OSA Journal of Lightwave Technology*, vol. PP, no. 99, pp. 1-1, 2016.
- [43] H. Mardoyan *et al.*, "Single carrier 168-Gb/s Line-Rate PAM direct detection transmission using high-speed selector power DAC for optical interconnects," *IEEE/OSA Journal of Lightwave Technology*, vol. 34, no. 7, pp. 1593-1598, Apr. 2016.
- [44] E. El-Fiky, M. Chagnon, M. Sowailam, A. Samani, M. Morsy-Osman, and D. V. Plant, "168-Gb/s single carrier PAM4 transmission for intra-data center optical interconnects," *IEEE Photonics Technology Letters*, vol. 29, no. 3, pp. 314-317, Feb. 2017.
- [45] A. S. Karar and J. C. Cartledge, "Generation and detection of a 56 Gb/s signal using a DML and half-cycle 16-QAM Nyquist-SCM," *IEEE Photonics Technology Letters*, vol. 25, no. 8, pp. 757-760, Apr. 2013.

- [46] K. Zhong *et al.*, “Experimental demonstration of 608Gbit/s short reach transmission employing half-cycle 16QAM Nyquist-SCM signal and direct detection with 25Gbps EML,” *OSA Optics express*, vol. 24, no. 22, pp. 25057-25067, Oct. 2016.
- [47] J. L. Wei, J. D. Ingham, D. G. Cunningham, R. V. Penty, and I. H. White, “Performance and power dissipation comparisons between 28 Gb/s NRZ, PAM, CAP and optical OFDM systems for data communication applications,” *IEEE/OSA Journal of Lightwave Technology*, vol. 30, no. 20, pp. 3273-3280, Oct. 2012.
- [48] J. L. Wei, D. G. Cunningham, R. V. Penty, and I. H. White, “Study of 100 Gigabit Ethernet using carrierless amplitude/phase modulation and optical OFDM,” *IEEE/OSA Journal of Lightwave Technology*, vol. 31, no. 9, pp. 1367-1373, May 2013.
- [49] M. I. Olmedo *et al.*, “Multiband carrierless amplitude phase modulation for high capacity optical data links,” *IEEE/OSA Journal of Lightwave Technology*, vol. 32, no. 4, pp. 798-804, Feb. 2014.
- [50] T. Tanaka *et al.*, “Experimental demonstration of 448-Gbps+ DMT transmission over 30-km SMF,” in *Proc. OFC 2014*, San Francisco, CA, 9-13 Mar. 2014, p. M2I.5.
- [51] W. Yan *et al.*, “80 km IM-DD transmission for 100 Gb/s per lane enabled by DMT and nonlinearity management,” in *Proc. OFC 2014*, San Francisco, CA, 9-13 Mar. 2014, p. M2I.4.
- [52] S. Chandrasekhar, X. Liu, P. J. Winzer, J. Simsarian, and R. Griffin, “Compact all-InP laser-vector-modulator for generation and transmission of 100-Gb/s PDM-QPSK and 200-Gb/s PDM-16-QAM,” *IEEE/OSA Journal of Lightwave Technology*, vol. 32, no. 4, pp. 736-742, Feb. 2014.
- [53] T. Tatsumi *et al.*, “A compact low-power 224-Gb/s DP-16QAM modulator module with InP-based modulator and linear driver ICs,” in *Proc. OFC 2014*, San Francisco, CA, 9-13 Mar. 2014, p. Tu3H.5.
- [54] M. Morsy-Osman *et al.*, “Colorless and preamplifierless reception using an integrated Si-photonics coherent receiver,” *IEEE Photonics Technology Letters*, vol. 25, no. 11, pp. 1027-1030, June 2013.
- [55] C. R. Doerr *et al.*, “Packaged monolithic silicon 112-Gb/s coherent receiver,” *IEEE Photonics Technology Letters*, vol. 23, no. 12, pp. 762-764, June 2011.

- [56] P. de Dobbelaere *et al.*, "Silicon photonics for high data rate optical interconnect," in *Proc. OIC 2012*, Santa Fe, NM, 20-23 May 2012, p. WA2.
- [57] Y. Painchaud *et al.*, "Ultra-compact coherent receiver based on hybrid integration on silicon," in *Proc. OFC 2013*, Anaheim, CA, 17-21 Mar. 2013, p. OM2J.2.
- [58] M. Y. S. Sowailem *et al.*, "100G and 200G single carrier transmission over 2880 and 320 km using an InP IQ modulator and Stokes vector receiver," *OSA Optics Express*, vol. 24, no. 26, pp. 30485-30493, Dec. 2016.
- [59] M. Sowailem *et al.*, "400-G single carrier 500-km transmission with an InP dual polarization IQ modulator," *IEEE Photonics Technology Letters*, vol. 28, no. 11, pp. 1213-1216, June 2016.
- [60] M. Y. S. Sowailem *et al.*, "770 Gb/s PDM-32QAM coherent transmission using InP dual polarization IQ modulator," *IEEE Photonics Technology Letters*, vol. 29, no. 5, pp. 442-445, Mar. 2017.
- [61] I. G. Lopez *et al.*, "DAC-free ultra-low-power dual-polarization 64-QAM transmission at 32 GBd with hybrid InP IQ SEMZM and BiCMOS drivers module," *IEEE/OSA Journal of Lightwave Technology*, vol. PP, no. 99, pp. 1-1, 2016.
- [62] N. Kikuchi, R. Hirai, and Y. Wakayama, "High-speed optical 64QAM signal generation using InP-based semiconductor IQ modulator," in *Proc. OFC 2014*, San Francisco, CA, 9-13 Mar. 2014, p. M2A.2.
- [63] Y. Ogiso *et al.*, "Ultra-high bandwidth InP IQ modulator with 1.5 V V_{π} ," in *Proc. ECOC 2016*, Düsseldorf, Germany, 18-22 Sept. 2016, pp. 355-357.
- [64] G. Letal *et al.*, "Low loss InP C-band IQ modulator with 40GHz bandwidth and 1.5 V V_{π} ," in *Proc. OFC 2015*, Los Angeles, CA, 22-26 Mar. 2015, p. Th4E.3.
- [65] J. G. Proakis, *Digital communications*, 4th ed. Boston: McGraw-Hill, 2001.
- [66] M. S. Faruk and K. Kikuchi, "Compensation for in-phase/quadrature imbalance in coherent-receiver front end for optical quadrature amplitude modulation," *IEEE Photonics Journal*, vol. 5, no. 2, pp. 7800110-7800110, Apr. 2013.
- [67] P. Ciblat and M. Ghogho, "Blind NLLS carrier frequency-offset estimation for QAM, PSK, and PAM modulations: performance at low SNR," *IEEE Transactions on Communications*, vol. 54, no. 10, pp. 1725-1730, Oct. 2006.

- [68] Y. Wang, E. Serpedin, and P. Ciblat, "An alternative blind feedforward symbol timing estimator using two samples per symbol," *IEEE transactions on communications*, vol. 51, no. 9, pp. 1451-1455, Sept. 2003.
- [69] I. Fatadin, D. Ives, and S. J. Savory, "Blind equalization and carrier phase recovery in a 16-QAM optical coherent system," *IEEE/OSA Journal of Lightwave Technology*, vol. 27, no. 15, pp. 3042-3049, Aug. 2009.
- [70] I. Fatadin and S. J. Savory, "Compensation of frequency offset for 16-QAM optical coherent systems using QPSK partitioning," *IEEE Photonics Technology Letters*, vol. 23, no. 17, pp. 1246-1248, Sept. 2011.
- [71] Y. Han and G. Li, "Coherent optical communication using polarization multiple-input-multiple-output," *OSA Optics Express*, vol. 13, no. 19, pp. 7527-7534, Sept. 2005.
- [72] E. Ip and J. M. Kahn, "Digital equalization of chromatic dispersion and polarization mode dispersion," *IEEE/OSA Journal of Lightwave Technology*, vol. 25, no. 8, pp. 2033-2043, Aug. 2007.
- [73] K. Kikuchi, "Fundamentals of coherent optical fiber communications," *IEEE/OSA Journal of Lightwave Technology*, vol. 34, no. 1, pp. 157-179, Jan. 2016.
- [74] M. Kuschnerov *et al.*, "DSP for coherent single-carrier receivers," *IEEE/OSA Journal of Lightwave Technology*, vol. 27, no. 16, pp. 3614-3622, Aug. 2009.
- [75] S. J. Savory, "Digital coherent optical receivers: algorithms and subsystems," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 16, no. 5, pp. 1164-1179, Oct. 2010.
- [76] S. J. Savory, "Digital filters for coherent optical receivers," *OSA Optics Express*, vol. 16, no. 2, pp. 804-817, Jan. 2008.
- [77] S. Benedetto and P. Poggiolini, "Theory of polarization shift keying modulation," *IEEE Transactions on Communications*, vol. 40, no. 4, pp. 708-721, Apr. 1992.
- [78] M. Chagnon, M. Morsy-Osman, D. Patel, V. Veerasubramanian, A. Samani, and D. Plant, "Digital signal processing for dual-polarization intensity and interpolarization phase modulation formats using stokes detection," *IEEE/OSA Journal of Lightwave Technology*, vol. 34, no. 1, pp. 188-195, Jan. 2016.

- [79] D. Che, Q. Hu, X. Chen, and W. Shieh, "1-Tb/s Stokes vector direct detection over 480-km SSMF transmission," presented at *OECC 2014*, Melbourne, Australia, 6-10 July, 2014. PDP.
- [80] D. Che, A. Li, X. Chen, Q. Hu, Y. Wang, and W. Shieh, "Stokes vector direct detection for linear complex optical channels," *IEEE/OSA Journal of Lightwave Technology*, vol. 33, no. 3, pp. 678-684, Feb. 2015.
- [81] D. Che, A. Li, X. Chen, Q. Hu, Y. Wang, and W. Shieh, "Stokes vector direct detection for short-reach optical communication," *OSA Optics Letters*, vol. 39, no. 11, pp. 3110-3113, June 2014.
- [82] D. Che and W. Shieh, "Polarization demultiplexing for stokes vector direct detection," *IEEE/OSA Journal of Lightwave Technology*, vol. 34, no. 2, pp. 754-760, Jan. 2016.
- [83] M. Morsy-Osman, M. Chagnon, M. Poulin, S. Lessard, and D. V. Plant, "224-Gb/s 10-km transmission of PDM PAM-4 at 1.3 μm using a single intensity-modulated laser and a direct-detection MIMO DSP-based receiver," *IEEE/OSA Journal of Lightwave Technology*, vol. 33, no. 7, pp. 1417-1424, Apr. 2015.
- [84] M. Morsy-Osman, M. Chagnon, and D. V. Plant, "Four-dimensional modulation and Stokes direct detection of polarization division multiplexed intensities, inter polarization phase and inter polarization differential phase," *IEEE/OSA Journal of Lightwave Technology*, vol. 34, no. 7, pp. 1585-1592, Apr. 2016.
- [85] K. V. Vishwanath, A. Greenberg, and D. A. Reed, "Modular data centers: how to design them?," in *Proc. LSAP 2009*, Garching, Germany, 10 June 2009, pp. 3-10.
- [86] M. Al-Fares, A. Loukissas, and A. Vahdat, "A scalable, commodity data center network architecture," *ACM SIGCOMM Computer Communication Review*, vol. 38, no. 4, pp. 63-74, Aug. 2008.
- [87] G. I. Papadimitriou, C. Papazoglou, and A. S. Pomportsis, "Optical switching: switch fabrics, techniques, and architectures," *IEEE/OSA Journal of Lightwave Technology*, vol. 21, no. 2, pp. 384-405, Feb. 2003.
- [88] M. Glick, D. G. Andersen, M. Kaminsky, and L. Mummert, "Dynamically reconfigurable optical links for high-bandwidth data center networks," in *Proc. OFC 2009*, San Diego, CA, 22-26 Mar. 2009, p. OTuA3.

- [89] O. Liboiron-Ladouceur *et al.*, "The data vortex optical packet switched interconnection network," *IEEE/OSA Journal of Lightwave Technology*, vol. 26, no. 13, pp. 1777-1789, July 2008.
- [90] M. Sowailam, D. Plant, and O. Liboiron-Ladouceur, "Contention resolution strategy in optical burst switched datacenters," in *Proc. OFC 2013*, Anaheim, CA, 17-21 Mar. 2013, p. JW2A.77.
- [91] M. Karol, M. Hluchyj, and S. Morgan, "Input versus output queueing on a space-division packet switch," *IEEE Transactions on Communications*, vol. 35, no. 12, pp. 1347-1356, Dec. 1987.
- [92] N. McKeown, "The iSLIP scheduling algorithm for input-queued switches," *IEEE/ACM Transactions on Networking*, vol. 7, no. 2, pp. 188-201, Apr. 1999.
- [93] N. McKeown, A. Mekkittikul, V. Anantharam, and J. Walrand, "Achieving 100% throughput in an input-queued switch," *IEEE Transactions on Communications*, vol. 47, no. 8, pp. 1260-1267, Aug. 1999.
- [94] J. Y. Wei and R. I. McFarland, "Just-in-time signaling for WDM optical burst switching networks," *IEEE/OSA Journal of lightwave technology*, vol. 18, no. 12, pp. 2019-2037, Dec. 2000.
- [95] S. Choudhury, K. Paniker, A. Sinha, S. Mukhopadhyay, and N. C. Debnath, "A proactive burst assembly algorithm for optical burst switched networks," in *Proc. ISCC 2008*, Marrakech, 6-9 July 2008, pp. 543-548.
- [96] T. Benson, A. Anand, A. Akella, and M. Zhang, "Understanding data center traffic characteristics," *ACM SIGCOMM Computer Communication Review*, vol. 40, no. 1, pp. 92-99, Jan. 2010.
- [97] H. Wang, A. Wonfor, K. Williams, R. Penty, and I. White, "Demonstration of a lossless monolithic 16x16 QW SOA switch," in *Proc. ECOC 2009*, Vienna, 20-24 Sept. 2009, vol. 2009, pp. 1-2.
- [98] W. E. Leland, M. S. Taqqu, W. Willinger, and D. V. Wilson, "On the self-similar nature of Ethernet traffic (extended version)," *IEEE/ACM Transactions on networking*, vol. 2, no. 1, pp. 1-15, Feb. 1994.

- [99] S. Fernandes, C. Kamienski, and D. Sadok, "Accurate and fast replication on the generation of fractal network traffic using alternative probability models," in *Proc. ITCOM 2003*, Orlando, FL, 6 Aug. 2003, pp. 154-163.
- [100] M. S. Taqqu, V. Teverovsky, and W. Willinger, "Estimators for long-range dependence: an empirical study," *Fractals*, vol. 3, no. 04, pp. 785-798, 1995.
- [101] *Hurst parameter estimate*. Available: <http://www.mathworks.com/matlabcentral/fileexchange/19148-hurst-parameter-estimate>
- [102] *IEEE Standard for Information technology-- Local and metropolitan area networks-- Specific requirements-- Part 3: CSMA/CD Access Method and Physical Layer Specifications Amendment 4: Media Access Control Parameters, Physical Layers, and Management Parameters for 40 Gb/s and 100 Gb/s Operation*, 2010.
- [103] M. Chagnon *et al.*, "Experimental study of 112 Gb/s short reach transmission employing PAM formats and SiP intensity modulator at 1.3 μm ," *OSA Optics Express*, vol. 22, no. 17, pp. 21018-21036, Aug. 2014.
- [104] T. Chan, I. C. Lu, C. Jyehong, and W. I. Way, "400-Gb/s transmission over 10-km SSMF using discrete multitone and 1.3- μm EMLs," *IEEE Photonics Technology Letters*, vol. 26, no. 16, pp. 1657-1660, Aug. 2014.
- [105] G. Stepniak, "Comparison of efficiency of N -dimensional CAP modulations," *IEEE/OSA Journal of Lightwave Technology*, vol. 32, no. 14, pp. 2516-2523, July 2014.
- [106] J. Wei, Q. Cheng, D. G. Cunningham, R. V. Penty, and I. H. White, "100-Gb/s hybrid multiband (HMB) CAP/QAM signal transmission over a single wavelength," *IEEE/OSA Journal of Lightwave Technology*, vol. 33, no. 2, pp. 415-423, Jan. 2015.
- [107] P. Dong, X. Chen, K. Kim, S. Chandrasekhar, Y.-K. Chen, and J. H. Sinsky, "128-Gb/s 100-km transmission with direct detection using silicon photonic Stokes vector receiver and I/Q modulator," *OSA Optics Express*, vol. 24, no. 13, pp. 14208-14214, June 2016.
- [108] M. Y. Sowailem, M. Morsy-Osman, O. Liboiron-Ladouceur, and D. V. Plant, "A self-coherent system for short reach applications," in *Proc. Photonics North (PN) 2016*, Quebec City, QC, 24-26 May 2016, pp. 1-1.
- [109] B. J. Puttnam *et al.*, "High-capacity self-homodyne PDM-WDM-SDM transmission in a 19-core fiber," *OSA Optics Express*, vol. 22, no. 18, pp. 21185-21191, Sept. 2014.

- [110] B. J. Puttnam *et al.*, “Investigating self-homodyne coherent detection in a 19 channel space-division-multiplexed transmission link,” *OSA Optics Express*, vol. 21, no. 2, pp. 1561-1566, Jan. 2013.
- [111] W. D. Sacher, T. Barwicz, B. J. F. Taylor, and J. K. S. Poon, “Polarization rotator-splitters in standard active silicon photonics platforms,” *OSA Optics Express*, vol. 22, no. 4, pp. 3777-3786, Feb. 2014.
- [112] Q. Xu, L. Chen, M. G. Wood, P. Sun, and R. M. Reano, “Electrically tunable optical polarization rotation on a silicon chip using Berry's phase,” *Nat Commun*, vol. 5, p. 5337, Nov. 2014.
- [113] I. Fatadin, D. Ives, and S. J. Savory, “Compensation of frequency offset for differentially encoded 16- and 64-QAM in the presence of laser phase noise,” *IEEE Photonics Technology Letters*, vol. 22, no. 3, pp. 176-178, Feb. 2010.
- [114] W. Idler, F. Buchali, and K. Schuh, “Experimental study of symbol-rates and MQAM formats for single carrier 400 Gb/s and few carrier 1 Tb/s options,” in *Proc. OFC 2016*, Anaheim, CA, 20-24 Mar. 2016, p. Tu3A.7.
- [115] J. Zhang, J. Yu, Z. Jia, and H.-C. Chien, “400 G transmission of super-Nyquist-filtered signal based on single-carrier 110-GBaud PDM QPSK with 100-GHz grid,” *IEEE/OSA Journal of Lightwave Technology*, vol. 32, no. 19, pp. 3239-3246, Oct. 2014.
- [116] J. H. Ke, Y. Gao, and J. C. Cartledge, “400 Gbit/s single-carrier and 1 Tbit/s three-carrier superchannel signals using dual polarization 16-QAM with look-up table correction and optical pulse shaping,” *OSA Optics Express*, vol. 22, no. 1, pp. 71-84, Jan. 2014.
- [117] R. Rios-Müller *et al.*, “Spectrally-efficient 400-Gb/s single carrier transport over 7,200 km,” *IEEE/OSA Journal of Lightwave Technology*, vol. 33, no. 7, pp. 1402-1407, Apr. 2015.
- [118] F. Buchali, A. Klekamp, L. Schmalen, and D. Tomislav, “Implementation of 64QAM at 42.66 GBaud using 1.5 samples per symbol DAC and demonstration of up to 300 km fiber transmission,” in *Proc. OFC 2014*, San Francisco, CA, 9-13 Mar. 2014, p. M2A.1.
- [119] T. J. Xia *et al.*, “Transmission of 400G PM-16QAM channels over long-haul distance with commercial all-distributed Raman amplification system and aged standard SMF in field,” in *Proc. OFC 2014*, San Francisco, CA, 9-13 Mar. 2014, p. Tu2B.1.

- [120] G. Raybon *et al.*, “Single-carrier 400G interface and 10-channel WDM transmission over 4,800 km using all-ETDM 107-Gbaud PDM-QPSK,” in *Proc. OFC 2013*, Anaheim, CA, 17-21 Mar. 2013, p. PDP5A.5.
- [121] S. Randel *et al.*, “All-electronic flexibly programmable 864-Gb/s single-carrier PDM-64-QAM,” in *Proc. OFC 2014*, San Francisco, CA, 9-13 Mar. 2014, p. Th5C.8.
- [122] J. Zhang, J. Yu, B. Zhu, and H. C. Chien, “WDM transmission of single-carrier 120-GBaud ETDM PDM-16QAM signals over 1200-km terrestrial fiber links,” *IEEE/OSA Journal of Lightwave Technology*, vol. PP, no. 99, pp. 1-1, 2017.
- [123] T. Rahman *et al.*, “38.4Tb/s transmission of single-carrier serial line-rate 400Gb/s PM-64QAM over 328km for metro and data center interconnect applications,” in *Proc. OFC 2016*, Anaheim, CA, 20-24 Mar. 2016, p. W3G.1.
- [124] L. Dou *et al.*, “420Gbit/s DP-64QAM Nyquist-FDM single-carrier system,” in *Proc. OFC 2016*, 20-24 Mar. 2016, p. Tu3A.5.
- [125] V. Lal *et al.*, “Full C-band tunable coherent transmitter and receiver InP photonic integrated circuits,” in *Proc. ECOC 2016*, Düsseldorf, Germany, 18-22 Sept. 2016, pp. PDP 10-12.
- [126] M. Paskov, D. Lavery, and S. J. Savory, “Blind equalization of receiver in-phase/quadrature skew in the presence of Nyquist filtering,” *IEEE Photonics Technology Letters*, vol. 25, no. 24, pp. 2446-2449, Dec. 2013.
- [127] C. R. Fludger and T. Kupfer, “Transmitter impairment mitigation and monitoring for high baud-rate, high order modulation systems,” in *Proc. ECOC 2016*, Düsseldorf, Germany, 18-22 Sept. 2016, pp. 256-258.
- [128] H. Yamazaki, T. Goh, T. Hashimoto, A. Sano, and Y. Miyamoto, “Generation of 448-Gbps OTDM-PDM-16QAM signal with an integrated modulator using orthogonal CSRZ pulse,” in *Proc. OFC 2015*, Los Angeles, CA, 22-26 Mar. 2015, p. Th2A.18.
- [129] J. K. Fischer *et al.*, “8×448-Gb/s WDM transmission of 56-GBd PDM 16-QAM OTDM signals over 250-km ultra-large effective area fiber,” *IEEE Photonics Technology Letters*, vol. 4, no. 23, pp. 239-241, Feb. 2011.
- [130] D. Rafique, A. Napoli, S. Calabro, and B. Spinnler, “Digital preemphasis in optical communication systems: On the DAC requirements for terabit transmission applications,” *IEEE/OSA Journal of Lightwave Technology*, vol. 32, no. 19, pp. 3247-3256, Oct. 2014.

- [131] D. Rafique, "Fiber nonlinearity compensation: Commercial applications and complexity analysis," *IEEE/OSA Journal of Lightwave Technology*, vol. 34, no. 2, pp. 544-553, Jan. 2016.
- [132] *OIF-FEC-100-01.0- 100G Forward Error Correction White Paper*. Available: http://www.oiforum.com/public/documents/OIF_FEC_100G-01.0.pdf
- [133] *OIF-Tech-Options-400G-01.0- Technology options for 400G Implementation*. Available: <http://www.oiforum.com/wp-content/uploads/OIF-Tech-Options-400G-01.0.pdf>
- [134] D. H. Sim, H. Kim, and Y. C. Chung, "Direct-detection receiver for polarization-division-multiplexed OOK signals," *IEEE Photonics Technology Letters*, vol. 27, no. 21, pp. 2238-2241, Nov. 2015.
- [135] K. Kikuchi and S. Kawakami, "Multi-level signaling in the Stokes space and its application to large-capacity optical communications," *OSA Optics Express*, vol. 22, no. 7, pp. 7374-7387, Apr. 2014.
- [136] S. Betti, F. Curti, G. D. Marchis, and E. Iannone, "Multilevel coherent optical system based on Stokes parameters modulation," *IEEE/OSA Journal of Lightwave Technology*, vol. 8, no. 7, pp. 1127-1136, July 1990.
- [137] F. Chang, K. Onohara, and T. Mizuochi, "Forward error correction for 100 G transport networks," *IEEE Communications Magazine*, vol. 48, no. 3, pp. S48-S55, Mar. 2010.
- [138] P. Johannisson, M. Sjodin, M. Karlsson, E. Tipsuwannakul, and P. Andrekson, "Cancellation of nonlinear phase distortion in self-homodyne coherent systems," *IEEE Photonics Technology Letters*, vol. 22, no. 11, pp. 802-804, June 2010.
- [139] T. Miyazaki, "Linewidth-tolerant QPSK homodyne transmission using a polarization-multiplexed pilot carrier," *IEEE Photonics Technology Letters*, vol. 18, no. 2, pp. 388-390, Jan. 2006.
- [140] J. P. Gordon and H. Kogelnik, "PMD fundamentals: Polarization mode dispersion in optical fibers," *Proceedings of the National Academy of Sciences*, vol. 97, no. 9, pp. 4541-4550, Apr. 2000.
- [141] P. J. Reyes-Iglesias, I. Molina-Fernández, A. Moscoso-Mártir, and A. Ortega-Moñux, "High-performance monolithically integrated 120° downconverter with relaxed hardware constraints," *OSA Optics Express*, vol. 20, no. 5, pp. 5725-5741, Feb. 2012.