

NEXT-GENERATION HIGH-CAPACITY COMMUNICATIONS WITH HIGH FLEXIBILITY, EFFICIENCY AND RELIABILITY

A Dissertation
Presented to
The Academic Faculty

by

Rui Zhang

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Electrical and Computer Engineering

Georgia Institute of Technology
August 2022

COPYRIGHT © 2022 BY RUI ZHANG

NEXT-GENERATION HIGH-CAPACITY COMMUNICATIONS WITH HIGH FLEXIBILITY, EFFICIENCY AND RELIABILITY

Approved by:

Dr. Xiaoli Ma, Advisor
School of Electrical and Computer
Engineering
Georgia Institute of Technology

Dr. Umakishore Ramachandran
School of Computer Science
Georgia Institute of Technology

Dr. John Barry
School of Electrical and Computer
Engineering
Georgia Institute of Technology

Dr. Winston I. Way
CTO, R&D
NeoPhotonics Corporation

Dr. Matthieu Bloch
School of Electrical and Computer
Engineering
Georgia Institute of Technology

Date Approved: May 09, 2022

To my family and friends.

ACKNOWLEDGEMENTS

I must begin by thanking my research advisor, Dr. Gee-Kung Chang, for guiding and supporting me. Dr. Chang has modeled all facets of my work, from the way I think to the way I present my works. During the campus visit in Georgia tech in 2017, it was clear to me that Dr. Chang was enthusiastic to research and supportive to students. He inspired me with great vision and gave me helpful advice on my career path. He always gave me independence to explore my own research interests. To me, he is not only a respectful teacher but also a close friend. His passion for life also influences me a lot. Besides work, he taught me to enjoy life and nature. He is an excellent photographer, and he was always happy to share the photos he took and travelling experiences to us. I have strived to learn as much as possible from him in Georgia tech and it was my great pleasure to work with him.

I must also thank my research co-advisor, Dr. Xiaoli Ma. Dr. Ma has guided and inspired me as an exemplary women researcher. She was willing to share her experience about life and help me regarding my career path. I have learned a lot from her.

I was very fortunate to collaborate with Dr. Winston I. Way at NeoPhotonics for four years. To me, Dr. Way is also my research advisor. We worked on many exciting topics on high-capacity data center communications, which becomes part of the thesis. Dr. Way is instrumental in building up my skillsets during my first year. He is very careful in conducting research and writing. I will always treasure the hours I spent with him in

discussing a lot of research details. At the beginning, I didn't put too much attention to the details. It was his comments and questions that challenged me to be a more careful researcher. I would like to thank him for giving me the opportunity to explore and learn about the photonics hardware and coherent optical systems.

Next, I want to thank my dissertation committee members for their time. Dr. John Barry has given me the best digital communication course, which lays the foundation on my research. Dr. Matthieu Bloch and Dr. Umakishore Ramachandran have provided valuable comments and suggestions, which helps to improve the quality of my work.

A big thank you is owed to my colleagues in the research group. It was my pleasure to be in this big family and be teammates with you. I would like to thank Dr. Feng Lu for helping me to start my first research project. Thanks to Dr. Daniel Guidotti, Dr. You-Wei Chen, Dr. Mu Xu, Dr. Chin-Wei Hsu, Dr. Xizi Tang, Dr. Shuyi Shen, Dr. Qi Zhou, Dr. Jiale He, Dr. Hyunwoo Cho, Dr. Siming Liu, Shuang Yao, Yahya Alfadhli, Shang-Jen Su, Shahmeer Omar, and many other graduate students, and staff members.

I was also fortunate to work with exceptional people over the course of industry internships and visits. Thank you to Dr. Anthony Ng'oma, Dr. Po-Tsung Boris Shih, Dr. Xinying Li, Dr. Ji Qing, and all the other research scientists in Corning during my visit in 2019. Thank you to Dr. Wen-Jr Jiang, Konstantin Kuzmin, Reggie Juluri, Dr. Yi Weng and all the other engineers when I did my internships at NeoPhotonics. In particular, I want to call out Dr. Wen-Jr Jiang, who has always been ready to lend an ear when I have technical questions. Thank you to Dr. Vincent Houtsma, Dr. Dora Van Veen, Dr. Noriaki Kaneda,

Dr. Yannick Lefevre, Dr. Amitkumar Mahadevan at Nokia Bell Labs for an awesome summer in 2020.

I also would like to thank my undergraduate thesis advisor and collaborators, Dr. Zhiping Zhou, Dr. Qingzhong Deng, and Bowen Bai at Peking University, for taking me into the adventure of research. I also must thank Dr. S.J. Ben Yoo and Dr. Gengchen Liu at UC Davis for great research collaboration during my undergraduate. It was the undergraduate research that motivated me to pursue Ph.D. study.

Finally, and most importantly, I thank my family for their unconditional support during these years. I would like to thank my parents for your love, raising me, educating me, and encouraging me to pursue my dreams. Thanks to my brother, who brought me lots of laughers and happiness during the childhood. Thanks to Wenlong Mou, my smart and gifted fiancé, who motivates me to become better through his energy in life and research. I appreciate the time that we explored new things together and I feel incredibly lucky to be with you these years!

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	iv
LIST OF TABLES	ix
LIST OF FIGURES	x
LIST OF SYMBOLS AND ABBREVIATIONS	xv
SUMMARY	xx
CHAPTER 1. Introduction	1
1.1 Heterogeneous Data Communication Network	1
1.2 Optical Fiber Communications	3
1.3 All-Spectra Fiber-Wireless Convergence	6
1.4 Challenges in Next-Generation Heterogeneous Communications	9
1.4.1 Probabilistic Shaping and Geometric Shaping to Enable More Flexibility	10
1.4.2 Trade-off Analysis for Inter Data Center Communications	13
1.4.3 Efficient System Innovation for Intra Data Center Communications	14
1.4.4 Data-Efficient Nonlinearity Correction Design for Broadband systems	16
1.4.5 Ultra-reliable mmW and FSO system design	18
CHAPTER 2. Flexible Shaping Under Diverse Channel Conditions	19
2.1 Definitions of Basic System Parameters	19
2.2 Entropy Allocation in Fiber-mmW Link	23
2.2.1 Principle of Operation	24
2.2.2 Experimental Setup	31
2.2.3 Computation Complexity and Parameter Setting	33
2.2.4 Experimental Results	34
2.2.5 Summary	37
2.3 Shaping in High-Speed Fiber-FSO Link	37
2.3.1 Principle of Operation	38
2.3.2 Experimental Setup	40
2.3.3 Experimental Results	41
2.3.4 Summary	44
2.4 Shaping for Flexible Rate PON	44
2.4.1 Principle of Operation	45
2.4.2 Simulation Setup	47
2.4.3 Simulation Results with Flexible FEC Code Rates	47
2.4.4 Simulation Results with a Fixed FEC Code Rate	49
2.4.5 Summary	51
CHAPTER 3. Efficiency Analysis and Design for High-Capacity Links	53
3.1 Impact of Frequency Jitter and Linewidth on Coherent Systems	53

3.1.1	Theoretical Analysis	54
3.1.2	Experimental and Simulation Setup	61
3.1.3	Experimental and Simulation Results	62
3.1.4	Summary	67
3.2	Impact of Laser Flicker Noise on Ultra-Baud Rate Coherent Systems	68
3.2.1	Experimental Setup	68
3.2.2	Experimental Results	70
3.2.3	Summary	76
3.3	Cost-Efficient Coherent System Design	77
3.3.1	Experimental Setup	77
3.3.2	Experimental and Simulation Results	81
3.3.3	Summary	86
3.4	Data-Efficient Nonlinearity Correction Algorithm	86
3.4.1	Principle of Operation	86
3.4.2	Experiment Setup	92
3.4.3	Experimental Results	93
3.4.4	Summary	103
CHAPTER 4.	Reliable Fiber-Wireless System Design	104
4.1	Integrated mmW and FSO A-RoF Link	104
4.1.1	Principle of Operation	104
4.1.2	Experimental Setup	109
4.1.3	Experimental Results	110
4.1.4	Summary	116
CHAPTER 5.	Conclusion	117
5.1	Flexibility under Diverse Channel Conditions	117
5.1.1	Technical Contributions	117
5.1.2	Future Directions	119
5.2	Efficiency Analysis and Efficient System Design	120
5.2.1	Technical Contributions	120
5.2.2	Future Directions	122
5.3	Reliability in Fiber-Wireless Integrated Links	123
5.3.1	Technical Contributions	123
5.3.2	Future Directions	123
APPENDIX A.	Derivation of Laser Phase Noise Impact	125
A.1	Derivation of I_1	125
A.2	Derivation of $I_2 - I_1$	129
REFERENCES		131
VITA		142

LIST OF TABLES

Table 2. 1	SE of different signals with different OHs. ©IEEE 2020	32
Table 2. 2	SOA gain of PS+GS PAM8 with pairwise versus ROP ©IEEE 2020	49

LIST OF FIGURES

Figure 1. 1	State-of-art heterogeneous data communication network.	2
Figure 1. 2	System architecture of (a) IM-DD and (b) coherent systems.	3
Figure 1. 3	Current trend and applications of IM-DD and coherent systems.	5
Figure 1. 4	Fiber wireless integration and networking for all-spectra access network. ©IEEE 2020.	7
Figure 1. 5	Shaping enables gradual entropy adaptation under different SNRs.	10
Figure 1. 6	Shaping block diagram at the transmitter.	10
Figure 1. 7	Combination with parity bits for bipolar distribution and unipolar distribution. © OSA 2021	11
Figure 2. 1	Block diagram of the PAS scheme. Probability distribution of the amplitude part of (a) PS 64QAM, and (b) PS 256QAM. 2-D probability distribution of the assembled symbols of (c) PS 64QAM, and (d) PS 256QAM. © IEEE 2020	19
Figure 2. 2	PS unit-based entropy allocation. (a) Uniform M-QAM or PS with no entropy allocation. (b) Entropy allocation based on PS unit. ©IEEE 2020	24
Figure 2. 3	The 64QAM constellation diagram of the inner, corner and edge symbols in modeling. ©IEEE 2020	25
Figure 2. 4	Entropy allocation algorithm. (a) Modeling based algorithm for entropy allocation. (b) Chart flow for calculating P_e given SNR and entropy. ©IEEE 2020	28
Figure 2. 5	Experimental setup and DSP diagram of the fiber-mmW system with entropy allocation. ©IEEE 2020	31
Figure 2. 6	Convergence speed of the entropy allocation algorithm. ©IEEE 2020	34
Figure 2. 7	SNR or Assigned entropy over frequency. ©IEEE 2020	35
Figure 2. 8	Experimental results of PS 64QAM. ©IEEE 2020	36

Figure 2. 9	Experimental results of PS 256QAM. ©IEEE 2020	37
Figure 2. 10	Principle of operations of the PW PS PAM signal. ©OSA 2021	39
Figure 2. 11	Experimental setup of the fiber-FSO link based on unipolar distribution. ©OSA 2021	40
Figure 2. 12	Sweep peak-to-peak amplitude. © OSA 2021	42
Figure 2. 13	Results with fixed FEC code rate and the same information rate. © OSA 2021	43
Figure 2. 14	Results with the same spectral efficiency and flexible FEC code rate. ©OSA 2021	43
Figure 2. 15	Illustration of the investigated schemes using PAM4 as an example. © IEEE 2020	46
Figure 2. 16	Simulation setup of the flexible rate PON. ©IEEE 2020	47
Figure 2. 17	Simulation results of signals without pairwise constraint. ©IEEE 2020	48
Figure 2. 18	Simulation results of signals with pairwise constraint. ©IEEE 2020	48
Figure 2. 19	GMI versus NGMI threshold Rth with pairwise constraint. © IEEE 2020	50
Figure 2. 20	Spectral efficiency versus required ROP with practical LDPC. ©IEEE 2020	51
Figure 3. 1	Measured FN PSD in a coherent pluggable module. ©IEEE 2019	53
Figure 3. 2	Block diagram of the theoretical analysis with jitter tone. ©IEEE 2019	54
Figure 3. 3	Simulation setup to investigate jitter tone tolerance. ©IEEE 2019	61
Figure 3. 4	Experimental setup to investigate jitter tone tolerance. ©IEEE 2019	62
Figure 3. 5	Experimental and simulation results with the impact of jitter tone. (rec. opt. power= -8dBm). ©IEEE 2019	63

Figure 3. 6	Simulation results (for 64Gbaud/DP-16QAM signals) with single jitter sinusoidal tone to validate theoretical conclusion. ©IEEE 2019	65
Figure 3. 7	Simulation results of the impact of multiple tones. ©IEEE 2019	66
Figure 3. 8	Experimental setup for flicker noise investigation. ©OSA 2020	69
Figure 3. 9	FN PSD of different lasers in the experiment. ©OSA 2020	70
Figure 3. 10	BER versus OSNR of 16QAM signal. ©OSA 2020	71
Figure 3. 11	BER versus OSNR of 64QAM signal (a)in BtB case and (b) with 450-km transmission. ©OSA 2020	71
Figure 3. 12	Estimated phase fluctuations versus time duration with different effective linewidths. ©OSA 2020	73
Figure 3. 13	Experimental results of 96-Gbaud 32QAM signal. ©OSA 2020	74
Figure 3. 14	Cycle slip rate versus OSNR of 96-Gbaud 32QAM signal (a) in BtB case and (b) with 400km fiber. ©OSA 2020	75
Figure 3. 15	Experimental setup and DSP block diagrams of the SHD coherent link. (i) is used in amplified case. ©OSA 2022	78
Figure 3. 16	Polarization tracking scheme in SHD coherent system.	80
Figure 3. 17	Experimental results (link budget) of umamplified link. ©OSA 2022	82
Figure 3. 18	(a) BER vs ROP and (b) BER versus OSNR (LO = 3dBm, signal ROP=-0.5dBm) of amplified link. ©OSA 2022	83
Figure 3. 19	BER vs path mismatch length and de-rotation period (OSNR = 37.5dB, LO=3dBm, signal ROP = -0.5dBm). ©OSA 2022	84
Figure 3. 20	Phase variation over time of the captured signal.	84
Figure 3. 21	BER versus fiber length with 0m path mismatch. ©OSA 2022	85
Figure 3. 22	Semi-supervised VNLE with Lasso using SD and HD. ©IEEE 2021	89

Figure 3. 23	Experiment setup of the fiber-FSO converged link. ©IEEE 2021	92
Figure 3. 24	BER versus weight α for 50-Gbaud PAM4 (5% labeled data and 60% unlabeled data). ©IEEE 2021	94
Figure 3. 25	BER versus weight α for 35-Gbaud PAM8 (5% labeled data and 60% unlabeled data). ©IEEE 2021	94
Figure 3. 26	Optimal BER versus unlabeled data ratio with 5% labeled data for 50-Gbaud PAM4 and 35-Gbaud PAM8. ©IEEE 2021	95
Figure 3. 27	Optimal α and λ versus unlabeled data ratio with 5% labeled data for PAM8. ©IEEE 2021	96
Figure 3. 28	BER of 50-Gbaud PAM4 versus memory lengths using 3% labeled data and 60% unlabeled data. ©IEEE 2021	97
Figure 3. 29	BER versus V_{pp} and signal eye diagrams using 5% labeled data. ©IEEE 2020	98
Figure 3. 30	BER versus labeled data ratio using 60% unlabeled data. ©IEEE 2020	99
Figure 3. 31	BER versus ROP using 5% labeled data in FSO transmission. ©IEEE 2020	100
Figure 3. 32	BER versus line rates using 5% labeled data. ©IEEE 2020	101
Figure 3. 33	Required ROP versus line rates using 5% labeled data at the 1.22×10^{-2} threshold. ©IEEE 2020	102
Figure 4. 1	Working principles of the coordinated mapping. ©IEEE 2018	105
Figure 4. 2	Signal with CMC in the frequency and time domains. ©IEEE 2018	107
Figure 4. 3	Experimental setup of the mmW/FSO integrated RoF system. ©IEEE 2018	110
Figure 4. 4	Experimental results of mmW links in (a) BtB experiment and (b) after 21-km SMF. ©IEEE 2018	111
Figure 4. 5	Experimental results of FSO links. (a) EVM performance versus mapping ratio. Power sensitivity performance for three mapping ratios (b) BtB measurements and (c) after 21-km SMF. ©IEEE 2018	111

Figure 4. 6	Experimental results of the hybrid mmW and FSO RoF link. ©IEEE 2018	114
Figure 4. 7	Transmitted OFDM signal and recovered QAM signal with frequency/burst interference. ©IEEE 2018	115

LIST OF SYMBOLS AND ABBREVIATIONS

PON	passive optical network
IM-DD	intensity modulation and direct detection
mmW	millimeter wave
FSO	free space optics
PS	probabilistic shaping
SNR	signal to noise ratio
GS	geometric shaping
MB	Maxwell-Boltzmann
ROP	received optical power
EEPON	equalization-enhanced phase noise
DSP	digital signal processing
SDM-SHD	space-division multiplexing-based self-homodyne detection system
Lasso	least absolute shrinkage and selection operator
CMC	coordinated mapping and combining
CD	chromatic dispersion
CU	centralized unit
DU	distributed unit
C band	conventional band
EDFA	erbium-doped fiber amplifier
WDM	wavelength-division multiplexing
LD	laser diode
DAC	digital-to-analog convertor

PD	photodiode
ADC	analog-to-digital convertor
DP	dual-polarization
MZ	Mach-Zehnder
LO	local oscillator
BPD	balanced photodiode
PBS	polarization beam splitter
MIMO	multiple-input and multiple-output
AE	adaptive equalizer
ISI	inter symbol interference
TL	tunable laser
CPR	carrier phase recovery
CFO	constant frequency offset
UE	user equipment
RF	radio frequency
RoF	radio over fiber
OOK	on-off keying
PAM	pulse amplitude modulation
CCDM	constant composition distribution matching
DM	distribution matching
FEC	forward error correction code
PAS	probabilistic amplitude shaping
MSB	the most significant bit
LSB	the least significant bit
PW	pairwise

EVM	error vector magnitude
ECL	external cavity laser
DBR	distributed Bragg reflector
FN PSD	frequency noise power spectral density
DFB	distributed feedback
VNLE	Volterra nonlinear equalizer
NN	neural network
OLS	ordinary least square
SD	soft decision
HD	hard decision
ADCT	adaptive-diversity-combining technique
GMI	generalized mutual information
PRBS	pseudorandom binary sequence
QAM	quadrature amplitude modulation
NGMI	normalized generalized mutual information
AWGN	additive white Gaussian noise
OFDM	orthogonal frequency-division multiplexing
MAP	maximum a posteriori
ML	maximum likelihood
SER	symbol error rate
BER	bit error rate
KL	Kullback-Leibler
IFFT	inverse fast Fourier transform
CP	Cyclic Prefix
IF	intermediate frequency

ENOB	effective number of bits
DML	directly modulated laser
SMF	single-mode fiber
PA	power amplifier
ED	envelope detector
FEC OH	forward error correction code overhead
SE	spectral efficiency
IR	Information rate
RMS	root mean square
LDPC	low-density parity-check code
AC	alternating current
DC	direct current
ER	extinction ratio
AWG	arbitrary waveform generator
LLR	loglikelihood ratio
APD	avalanche photodiodes
E/O	electrical to optical
SOA	semiconductor optical amplifier
RIN	relative intensity noise
Tx	transmitter
Rx	receiver
BtB	back-to-back
ASE	amplified spontaneous emission
OSNR	optical signal-to-noise ratio
CDM	coherent driver modulator

ICR	intradynne coherent receiver
BPS	blind phase search
SHD	self-homodyne
ATT	attenuator
MUX	multiplexer
DEMUX	de-multiplexer
OSA	optical spectrum analyzer
PC	polarization controller
SOP	state of polarization
EQ	equalizer
HWP	half wave plate
APC	automatic polarization controller
sVNLE	Simplified Volterra nonlinear equalizer
ADMM	alternating direction method of multipliers
cFEC	concatenated forward error correction code
oFEC	open forward error correction code
rROP	Required received optical power
MZM	Mach–Zehnder modulator
LNA	low-noise amplifier
LTE	Long-Term Evolution

SUMMARY

The data traffic has experienced an exponential growth recently from emerging bandwidth-consuming services. Heterogeneous optical and wireless communication architectures need to be developed to deliver multiple services in different user scenarios. Fiber optics have become the choice to scale out data center networks, support long-haul transmission (up to thousands of kilometers) and passive optical networks (PONs) due to its low insertion loss, security, and high data throughput. As for radio access networks, fiber optics cover most transmission distance while wireless communications provide last-mile delivery and connect to different end users with high mobility and flexibility.

Optical communication systems exhibit a dichotomy between intensity modulation and direct detection (IM-DD) and coherent detection, featuring low cost and ultra-high capacity, respectively. IM-DD systems are widely deployed in cost-sensitive access networks, PON, and intra-data center communications, where the transmission distance is short, or the capacity requirement is relatively low. In contrast, coherent systems are used in inter-data center communications, long-haul applications, where the fiber reach is long, and the capacity is high. As for wireless communication systems, the available spectrum in wireless communication is becoming congested in access network as broadband services continue to grow. Higher operation frequency such as millimeter wave (mmW), Terahertz, free space optics (FSO), and visible light communication have attracted a lot of research attentions. Different frequency bands have different advantages and disadvantages in terms of data rate, coverage range, as well as channel and hardware impairments. Therefore, the future wireless systems can never rely on single frequency band operation. Different

frequency bands can support different services based on their properties. Moreover, the fiber-wireless convergence and photonics-aided mmW systems are promising to support all-spectra access network with high capacity.

The objective of this dissertation is to address the flexibility, efficiency and reliability in high-capacity heterogeneous communication systems. Specifically, the dissertation focuses on the following topics.

Shaping is promising to enable more flexibility under diverse channel conditions. While conventional uniform distribution adjusts the signal entropy through a stepwise fashion, probabilistic shaping (PS) achieves a gradual and continuous adaptation of signal entropy under different signal to noise ratios (SNRs). On the other hand, geometric shaping (GS) can offer more flexibility in system optimization by shifting the symbol positions and enabling non-uniform symbol spacing. Most experimental demonstrations are for coherent systems so far, where the probability distribution is bipolar Maxwell-Boltzmann (MB) distribution. In the dissertation, we will experimentally investigate the shaping techniques to improve link budgets, and further extend them to more diverse and complicated scenarios, which result in more flexible systems and better network resource usage. The scenarios include 1) entropy allocation scheme under uneven frequency response for multi-carrier fiber-mmW system based on modeling result of MB distribution, 2) fiber-FSO link based on IM-DD scheme using unipolar pairwise distribution and nonlinear pre-distortion, and 3) flexible rate PON with a wide range of received optical powers (ROPs). Moreover, GS is proposed to be combined with PS to mitigate the penalty from the combination with parity bits in flexible rate PON.

Next, we perform efficiency analysis in inter-data center and long-haul communications. The phase noise from tunable laser induces not only phase variation but also equalization-enhanced phase noise (EEPN) for long-distance transmission. EEPN is hard to be mitigated through digital signal processing (DSP) because it incurs inter-symbol interference. In this dissertation, we will investigate and characterize the impact of the laser linewidth, jitter tones from circuit noise and the flicker noise from thermal tuning of distributed Bragg reflector lasers on beyond 400G coherent systems with different channel baud rates and fiber lengths through theoretical analysis, simulation, and experimental validation. The trade-off analysis results indicate the importance of setting up frequency noise power spectral density masks to qualify the transceiver laser design for various system performance requirements.

Besides efficiency analysis, we will also work on efficient system architecture and algorithm design. For cost-sensitive intra-data center link, the coherent system needs to be simplified with less DSP and cheap laser. The space-division multiplexing-based self-homodyne detection system (SDM-SHD) becomes a potential key technology, which has lower cost than conventional coherent system and higher throughput than IM-DD. In this dissertation, we will perform a detailed experimental investigation for future beyond 800G SDM-SHD. We will investigate the combined impact of various hardware impairments using various simplified DSP schemes. We have experimentally demonstrated that the amplified SDM-SHD can greatly enhance the link capacity in comparison to the un-amplified case. Thus, C-band SHD transceivers are very promising for future intra-data center applications. On the other hand, to improve the data efficiency of the nonlinearity correction algorithm in broadband communication systems, we leverage the semi-

supervised method and least absolute shrinkage and selection operator (Lasso) with an experimental validation in a fiber-FSO converged transport system. Experimental results in a fiber-FSO link validate that Lasso can reduce the required pilot symbol number by exploiting the sparsity of the tap coefficients. Semi-supervised method can further enhance the system performance without introducing additional overhead.

On the other hand, mmW and FSO are sensitive to different weather and atmospheric conditions. Adaptively combining the two complementary links can enhance the reliability of the hybrid wireless system under different weather conditions and interferences without feedback complexity. In this dissertation, we propose and experimentally demonstrate a coordinated mapping and combining (CMC) technique that enables the implementation of an ultra-reliable integrated mmW/FSO analog radio over fiber system. In the CMC design, the data blocks at the transmitter side are interleaved and repeated in both the frequency domain and the time domain, and then the duplicated blocks are combined with adaptive weight coefficients at the receiver side. The multiple-spectra operation exploits the abundant FSO and mmW bandwidth and shows superior performance in reliability and sensitivity compared to the conventional systems, even in extreme weather conditions and with strong burst interference.

CHAPTER 1. INTRODUCTION

1.1 Heterogeneous Data Communication Network

Recent years have witnessed the exponential growth of data traffic from emerging bandwidth-consuming services such as the mission critical Internet of Things, 4k/8k ultra-high-definition video streaming, social networking, and machine learning based applications. According to Ericsson mobility report [1], 5G networks will carry 62 percent of total mobile data traffic by 2027. Also, the fixed wireless access data traffic will grow almost 6 times in 2027, and the internet of things connections have experienced explosive growth. All of the data traffic running at the edge also drives the increasing needs for data center capabilities including higher capacity, lower latency and new network topology, etc [2]. Heterogeneous optical and wireless communication networks are pivotal to meet the ever-growing demand of high-throughput data traffic and diverse application scenarios [3]. Versatile methodologies and system interface architectures need to be developed to deliver multiple services in different user environments.

Figure 1. 1 shows the architecture of heterogeneous data communication network, which includes data center communication, fixed access services such as passive optical networks (PONs), and radio access networks. Fiber optics have become the choice to scale out data center networks and support long-haul transmission (up to thousands of kilometers) due to its low insertion loss, security, and high data throughput up to Tbits per second. Inter-data center communication denotes the link between different data centers, whose distance can be hundreds of kilometers. Thus, fiber chromatic dispersion (CD) and

nonlinearities are main impairments that degrade the system performances with long distances.

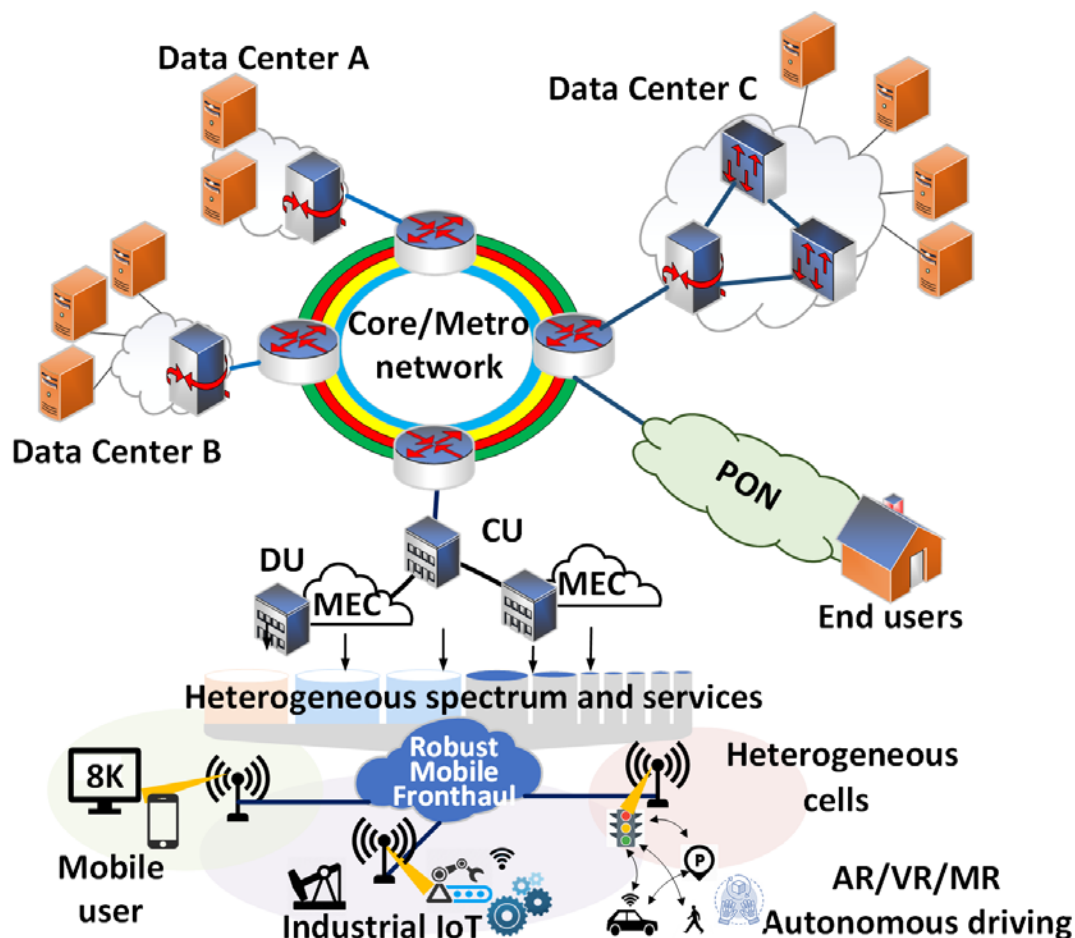


Figure 1. 1: State-of-art heterogeneous data communication network.

On the other hand, intra-data center communications support the data transmission within the data center, whose distance is usually shorter than two kilometers [2]. In such case, compared with inter-data center communications, intra-data center communications have more stringent requirements on cost efficiency, power consumption, latency, and density due to its sheer volume. Similarly, PON is also cost-sensitive and deploys optical fiber communications to support point-to-multiple point, high-speed (up to tens of Gbits

per second) fiber-to-the home services [4]. As for radio access networks, fiber optics cover most transmission distance of the mobile fronthaul (e.g., from the centralized unit (CU) to the distributed unit (DU)) while wireless communications can provide “last-mile” data delivery and connect to different end users with high mobility and flexibility [5].

1.2 Optical Fiber Communications

After the invention of laser in 1960 and pure silica fiber in 1970, the optical fiber communication becomes the reality. The typical operating wavelength bands span from 1260 nm to 1675 nm, which are between the range from 100 THz to 200 THz. In particular, the conventional band (C band) ranging from 1530nm to 1565nm is the primary wavelength band used for optical communication, especially in long-distance transmission, due to minimum attenuation and the availability of erbium-doped fiber amplifiers (EDFAs).

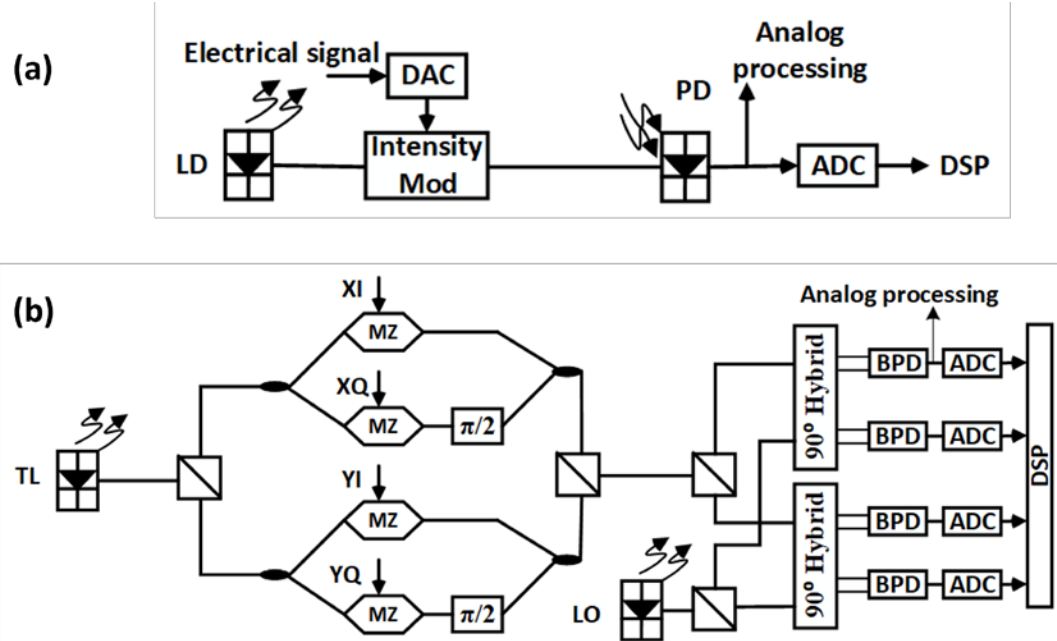


Figure 1. 2: System architecture of (a) IM-DD and (b) coherent systems.

Wavelength-division multiplexing (WDM) can further enhance the data throughput in optical fiber communication [6]. The use of multiplexers enables transmitting multiple wavelengths in a single fiber. Then the de-multiplexer at the receiver decouples different information carried through different wavelengths. The state-of-the-art dense WDM techniques use 40 channels with 100 GHz spacing or 80 channels with 50 GHz spacing in C band. Moreover, optical communication systems exhibit a dichotomy between intensity modulation and direct detection (IM-DD) and coherent detection, featuring low cost and ultra-high capacity, respectively. Figure 1. 2(a) shows the system architecture of IM-DD, which consists of a laser diode (LD), an intensity modulator and a digital-to-analog convertor (DAC) to convert the digital electrical signal to optical intensity. A photodiode (PD) detects the optical signal at the receiver side, which is followed by an analog-to-digital convertor (ADC), and digital signal processing (DSP) such as channel equalization and demodulation. The IM-DD scheme can only carry amplitude information, which has a low spectral efficiency. It only requires simple hardware structures and DSP featuring low costs. The dual-polarization (DP) coherent optical systems in Figure 1. 2(b) is much more complicated regarding both DSP algorithms and hardware. It modulates the signal in both the amplitude and the phase using several Mach-Zehnder (MZ) modulators and extends the signal to two orthogonal polarization states to double the link capacity. The receiver first splits the two polarizations using a polarization beam splitter (PBS), then down-converts and detects the signal using the local oscillator (LO), 90° optical hybrid and balanced PD (BPD). Non-ideal transceivers and polarization randomness in optical fibers introduce crosstalk between I and Q component as well as X and Y polarization in the coherent system. Thus, a multiple-input and multiple-output (MIMO) adaptive equalizer (AE) [7, 8]

is implemented to mitigate both inter symbol interference (ISI) from bandwidth limitation and the components crosstalk. Also, the independent transmitter tunable laser (TL) and LO introduces significant phase variation and frequency offset in the received signal. The receiver DSP needs to compensate for phase noises using carrier phase recovery (CPR) [9] and constant frequency offset (CFO) estimation [10, 11].

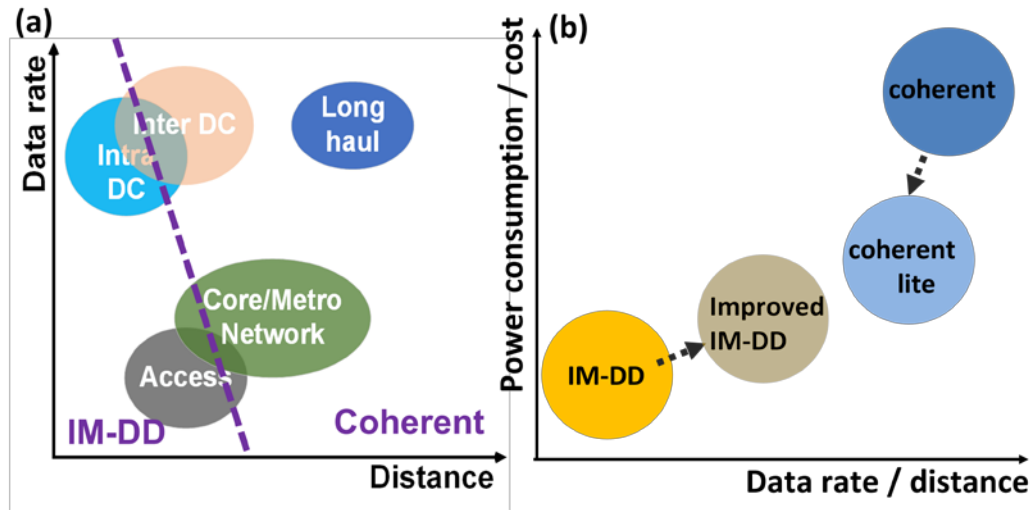


Figure 1. 3: Current trend and applications of IM-DD and coherent systems.

Figure 1. 3 summarizes the current trends of IM-DD and coherent systems. Coherent systems can support higher data rate or longer transmission distance than IM-DD systems due to higher spectral efficiency, polarization diversity, ease of CD compensation, and better sensitivity performance. However, due to the heavy DSP and more expensive devices, coherent systems have higher power consumptions and costs than IM-DD schemes. Thus, the IM-DD systems are widely deployed in cost-sensitive access networks, PON, and intra-data center communications, where the transmission distance is short, or the capacity requirement is relatively low. In contrast, coherent systems are used in inter-data center communications, long-haul applications, where the fiber reach is long, and the

capacity is high. Currently, as presented in Figure 1. 3(b), there are many research efforts on improving IM-DD systems such as improving the fiber dispersion tolerance and data throughput [12, 13]. On the other hand, coherent systems attract lots of attentions attempting to bring it to intra-data center communications [6, 14, 15] and PON in a cost-effective way, which is also named as coherent lite.

1.3 All-Spectra Fiber-Wireless Convergence

Compared with optical fiber communications, wireless communications are vital in connecting to end users with high mobility and more flexibility in access networks. The available spectrum in wireless communication is becoming congested in access network as the data traffic and broadband services continue to grow. Higher operation frequency such as millimeter wave (mmW), Terahertz, free space optics (FSO), and visible light communication have attracted a lot of research attentions [3, 16-18]. The hardware also migrates from electronics to photonics with the increase of operating frequency from sub-6GHz to FSO. Thus, different frequency bands show different advantages and disadvantages in terms of data rate, coverage range, as well as channel and hardware impairments. For instance, FSO has ultra-high capacity but shows low mobility and high system cost. The mmW system demonstrates good mobility and can support higher data rate than sub-6GHz, but its coverage is limited due to high attenuation. Visible light and light fidelity are more suitable to provide indoor data transmission and user equipment (UE) positioning. Therefore, the future wireless systems can never rely on single frequency band operation and designing a system combining the advantages of different spectra are feasible. Different frequency bands can serve as a complementary link [19, 20] or deal with

different users and services based on the functionality, cost, and coverage range, which is named as all-spectra operation.

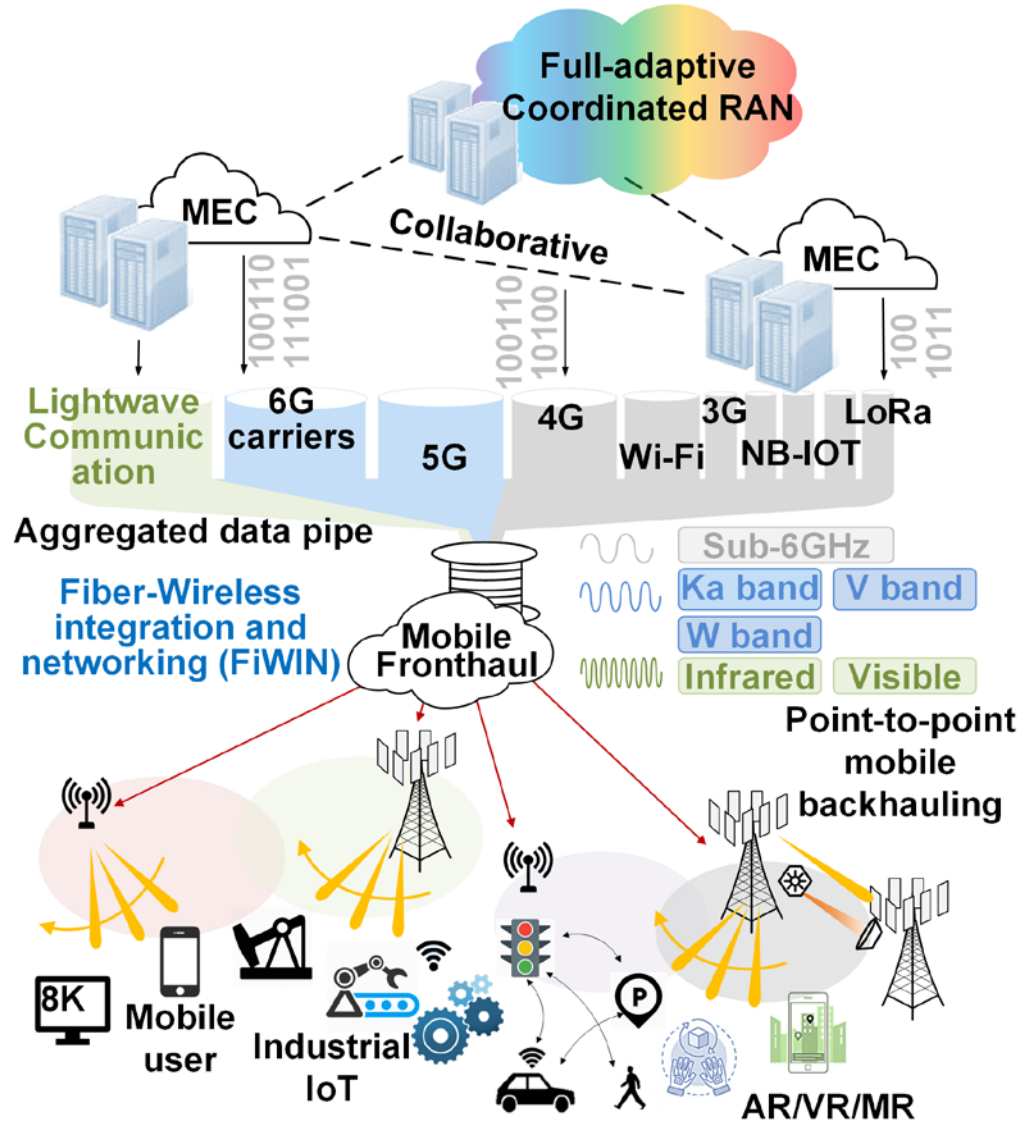


Figure 1. 4: Fiber wireless integration and networking for all-spectra access network. ©IEEE 2020.

Moreover, the fiber-wireless convergence is promising to support all-spectra access network with heterogeneous frequency bands, services, and cells. Figure 1. 4 demonstrates the concept of multiple spectra operations based on fiber-wireless integration. Various

carriers could be aggregated in the same fiber from the CU. The optical fiber can be easily scaled out through WDM to support massive connections. At the user end, different types of services are decoupled depending on their physical layer properties.

The combination of PD, electrical mixer, and frequency doubling/quadrupling can achieve a seamless and simple integration between the optical link and wireless link with a lower system bandwidth. To further conquer the bottleneck of low bandwidth operation in high-frequency electrical devices, photonics-assisted mmW or THz generation has been a mainstream technology [21, 22]. The two optical carriers with a certain frequency difference can be transmitted through the fiber and beat with each other at a PD. The PD directly up-converts them to radio frequency (RF) signal. The frequency difference between the two laser tones determines the carrier frequency of RF signal. Therefore, the scheme doesn't require mixer for up-conversion. Moreover, while digital radio over fiber (RoF) technique becomes mature, analog RoF has attracted massive research attentions to solve the bandwidth challenge and reduce the system latency. The photonics-assisted RF signal generation is also compatible with analog RoF. Analog RoF, however, is more sensitive to nonlinearity and noise than digital RoF.

On the other hand, PON is a last-mile, fiber-to-the-x telecommunications network that broadcasts data through fiber optic cables, which has been standardized as Gigabit Ethernet PON, 10G-PON, Next-Generation Passive Optical Network 2 [23]. Typically, PON employs IM-DD scheme with single carrier On-off keying (OOK) and Pulse Amplitude Modulation (PAM) modulations or discrete multitone. Due to its passive point to multiple point architecture, PON has the promise to transport all of these in a cost-effective way [24]. Also, existing PON networks already provide a substantial footprint

that operators can tap into. However, up till now PON standards have employed a fixed data rate designed for the worse-case channel conditions [23]. This results in sub-optimal resource utilization since different users see different channel conditions.

1.4 Challenges in Next-Generation Heterogeneous Communications

Next-generation communication systems face challenges in many aspects including capacity, flexibility, reliability, latency, efficiency, and coverage. The high-capacity communication systems employ high operation frequency bands such as mmW and FSO in wireless communications, and coherent detection in optical communications, respectively. This dissertation aims to address the flexibility, efficiency, and reliability in high-capacity communication systems under diverse application scenarios, including both wireless and optical communications.

Due to more diverse use scenarios, it is challenging to fully optimize the resource usage and data throughput with more flexibility under different and complicated channel conditions. The emerging techniques such as network slicing, machine learning based resource allocation, and probabilistic shaping under different signal to noise ratios (SNRs) enable a flexible communication network [25, 26]. On the other hand, some applications such as autonomous driving requires ultra-reliable data transmission. Interference cancellation and avoidance techniques and multiple-spectra operation have been used to enhance the system reliability [19, 27]. Moreover, while the system performance could be further improved through more complicated device components and DSP algorithms, the system power consumption and cost will be high. The trade-off analysis between the system cost and performance is critical for end-to-end optimization of communication

systems under different users. Joint hardware architecture and DSP design can further improve the system efficiency. Specifically, the dissertation focuses on the following topics.

1.4.1 Probabilistic Shaping and Geometric Shaping to Enable More Flexibility

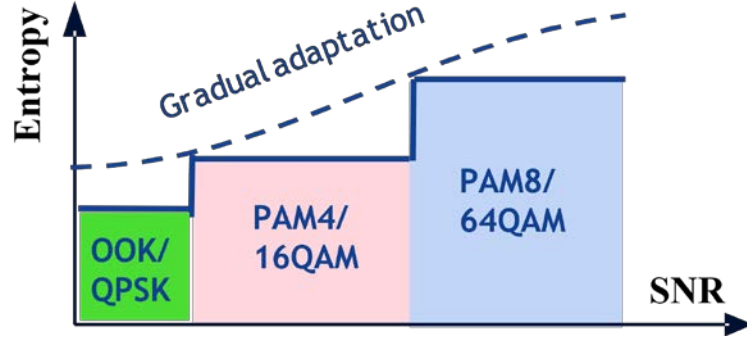


Figure 1. 5: Shaping enables gradual entropy adaptation under different SNRs.

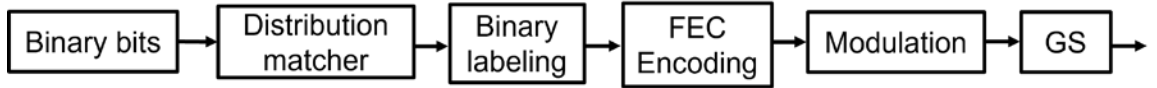


Figure 1. 6: Shaping block diagram at the transmitter.

From the information-theoretic point of view, Gaussian distribution or Maxwell-Boltzmann (MB) distribution can approach the Shannon capacity well when the average power constraint is determined by the second moment of signal amplitude [28]. However, shaping in real-world system is still not fully investigated. Recently, constant composition distribution matching (CCDM) based on arithmetic coding has been proposed to encode binary sequence with Bernoulli distribution to the symbol sequence with non-uniform distribution [29]. There are also other research efforts to reduce the distribution matching (DM) complexity and rate loss using lookup table, shell mapping etc [30-32]. The development of DM also drives the probabilistic shaping (PS) being deployed in optical

communication systems. As shown in Figure 1. 5, while uniform distribution adjusts the signal entropy through a stepwise fashion, shaping achieves a gradual and continuous adaptation of signal entropy. PS adaptively optimizes the mutual information and signal entropy under different SNRs by enabling non-uniform probability distributions for different symbol levels.

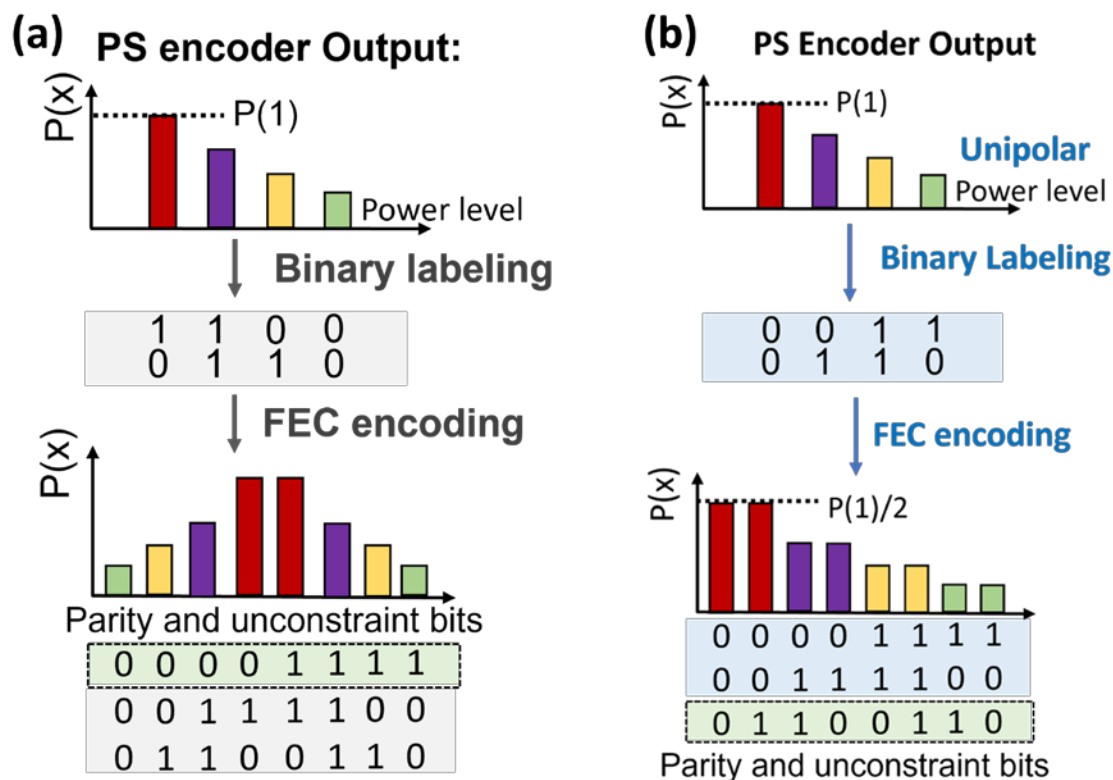


Figure 1. 7: Combination with parity bits for bipolar distribution and unipolar distribution. © OSA 2021

It is worth mentioning that the combination of forward error correction (FEC) codes and DM becomes non-trivial at the transmitter [26]. Figure 1. 6 illustrates the shaping block diagram at the transmitter, where the DM is put before FEC encoding. The shaped symbol sequence (output of the DM) is first mapped back to binary bits sequence. Then the binary

bits sequence is combined with FEC parity bits, which is followed by signal modulation. The conventional serially concatenated combination with parity bits results in a different probability distribution of the output symbol sequence, compared to the output of DM. Therefore, the combination of shaping with FEC parity bits need to be adjusted to preserve the probability distribution from DM. On the other hand, geometric shaping (GS) can offer more flexibility in system optimization by shifting the symbol positions and enabling non-uniform symbol spacing [33]. However, GS requires ADC with a high resolution.

Most experimental demonstrations are for coherent systems so far, where the probability distribution is bipolar MB distribution [34, 35]. Bipolar distribution can be combined with FEC using probabilistic amplitude shaping (PAS) [26] as shown in Figure 1. 7 (a). The DM generates half of the symbol sequence, which serves as the signal amplitude. Then the parity bits are formed as the most significant bit (MSB) and determine the signal sign. However, in a single-carrier IM-DD system with an optical amplifier, the optical power is determined by the first moment of the signal. In such case, the entropy-maximizing distribution is the unipolar exponential distribution [36, 37]. To preserve the unipolar trend, as shown in Figure 1. 7 (b), the parity bits are inserted into the least significant bit (LSB) to form a pairwise (PW) distribution [12]. The PW unipolar distribution has showed negligible degradation in certain SNR ranges.

In the dissertation, we will experimentally investigate the shaping techniques, and further extend it to more diverse and complicated scenarios. The investigation includes entropy allocation for multi-carrier fiber-mmW system [38-40], fiber-FSO link based on IM-DD scheme [41], and flexible rate PON with a wide range of ROPs [42, 43]. Moreover,

GS is proposed to be combined with PS to fully optimize the performance of flexible rate PON.

1.4.2 Trade-off Analysis for Inter Data Center Communications

Tunable lasers are critical to optical networking and coherent optical transmission, and their linewidths are key parameters which impact system performance [44, 45], especially for inter-data center communications and long-haul transmission. Laser white frequency noise can be modelled as a Wiener process and determines the laser intrinsic linewidth. It induces not only phase variation but also equalization-enhanced phase noise (EEPN) for long-distance transmission. Several studies have shown that EEPN is a conversion from phase to amplitude noise. It is caused by intermixing of the received dispersed signal, and the phase-noise-induced sidebands of the LO in a coherent system with post-reception electrical chromatic CD compensation [46, 47]. It was found that the square of error vector magnitude (EVM) due to EEPN is proportional to the product of linewidth, baud rate and fiber length [48]. Typically, there are three types of lasers with different intrinsic linewidths. The external cavity lasers (ECLs) have a narrow linewidth as low as tens of kHz. The distributed Bragg reflector (DBR) lasers exhibit linewidth of hundreds of kHz. Distributed feedback lasers (DFBs) have a large linewidth of several MHz with the lowest cost. In practice, ECL and DBR are two promising candidates for deployment in beyond 400G systems.

However, only considering the laser linewidth is not sufficient to characterize the system performance since there are other different phase noises in the system. The laser frequency noise (FN) power spectral density (PSD) can be used to characterize the phase

noise in the coherent system, where FN in different frequency region shows different impact on the system performance [49, 50]. The FN PSD is measured as follows. A well-designed, polarization-maintaining Mach-Zehnder interferometer serves as an optical frequency discriminator. The discriminator measures FN PSD by converting the frequency noise to intensity fluctuation, which is followed by intensity noise removing [51, 52]. The white FN PSD in high frequency region is modelled as a Wiener process, whose level is directly related to laser's Lorentzian 3-dB linewidth by a factor of π [53]. On the other hand, DBR laser has a much higher flicker noise compared to other types of lasers, which results in a high level in the low frequency of the FN PSD. It has been shown to cause significant coherent system penalty even without any transmission fiber [54, 55]. However, the examined flicker noise level was excessively high due to the current injection-induced carrier density fluctuation in the phase sections. This excessive low frequency noise can be reduced by using thermal tuning instead [56]. Also, the sinusoidal interfering tones, which are caused by switching power supplies, power converters, and other circuit noise in pluggable modules and line-cards [57-59], have not been carefully studied.

In this dissertation, we will investigate and characterize the impact of the laser linewidth, jitter tones and the flicker noise on beyond 400G coherent systems with different channel conditions [60-62]. The trade-off analysis results guide the transceiver laser design and qualification for different system performance requirements.

1.4.3 Efficient System Innovation for Intra Data Center Communications

The case is different in intra data center communication, where the fiber is short (< 2km) and the dispersion is not significant. IM-DD scheme is the state-of-the-art technique

due to its low power consumption and cost [2]. However, IM-DD faces challenges with the increase of data throughput, due to its low spectral efficiency and limited available wavelengths. The conventional coherent system shows high spectral efficiency with higher power consumption and cost than the IM-DD scheme. Thus, for cost-sensitive intra-data center link, the coherent system needs to be simplified with less DSP and cheap laser. Coherent lite thus becomes a potential key technology, which has lower cost than conventional coherent system and higher throughput than IM-DD [6, 14, 63]. In the space-division multiplexing-based self-homodyne detection system (SDM-SHD), the transmitter laser is split into two paths. One goes into the optical modulator for signal modulation and the other transmits to the receiver side as a remote LO [64]. Therefore, the phases of the remote LO and the signal align with each other if the distance of signal path and LO path match with each other. In such case, the phase noise in the received signal is significantly reduced. The receiver DSP can bypass two DSP function blocks (CFO estimation and CPR) and can also relax the laser linewidth requirement (e.g., DFB could be used) [63]. Moreover, MIMO equalizer can be simplified if the polarization de-multiplexing is performed in analog domain. The analog polarization tracking is feasible because of the short fiber length and slow changing speed of polarization.

Till now, the coherent lite system is still not fully investigated. Some of the works run the experiment with a low baud rate and low order QAM, which is less sensitive to quadrature error and CD [6, 14]. On the other hand, the full MIMO AE are still used in high baud rate demonstration, which makes the power consumption relatively high. Thus, in this dissertation, we will perform a detailed experimental investigation for future beyond 800G SDM-SHD [65]. We will preliminarily design a polarization tracking scheme in

signal path and investigate the combined impact of various hardware impairments using different simplified DSP schemes.

1.4.4 Data-Efficient Nonlinearity Correction Design for Broadband systems

Broadband systems such as FSO have more stringent power budget requirement due to large signal bandwidth and stricter SNR requirements. Typically, deeper modulation depth, larger signal amplitude, as well as higher amplifier gain could improve the receiver sensitivity performance at the expense of introducing nonlinearities from modulators and amplifiers [66, 67]. Moreover, high-order modulation format has better spectral efficiency but is less tolerant to nonlinearities. Thus, nonlinearity correction techniques are adopted to tackle with these nonlinear impairments.

Volterra nonlinear equalizer (VNLE) is one of the common DSP techniques used in optical communication systems to mitigate nonlinearities from modulation, amplifiers, and channel impairments [68]. It employs a polynomial regression and restores the signal by fitting and applying the inverse nonlinear transfer function incurred by both the linear and the nonlinear ISI [69]. A simplified VNLE is introduced to reduce the computational complexity by removing the interaction terms and only keeping the memory polynomial terms [67, 70]. Furthermore, there are research works that theoretically or experimentally investigate sparse Volterra, as well as supervised VNLE using Lasso and coefficient pruning [71-74]. The added l_0 or l_1 regularization term in VNLE using Lasso enforces the insignificant tap coefficients to be zero. It has demonstrated significant reductions in complexity using regularization in both PON and optical interconnect systems. On the other hand, supervised and semi-supervised neural network (NN) based nonlinear

equalizers have been investigated for self-interference cancellation and system nonlinear compensation [27, 75]. However, NN typically exhibits slower convergence rates and higher complexity compared to VNLE, as the later can be solved through convex optimization. The nonlinearity in a fiber-FSO access link mainly comes from modulation and amplifiers, which can be well approximated by polynomial regression. Thus, VNLE is a sufficiently enough technique to mitigate the nonlinearity. The Volterra nonlinear equalizer requires a large amount of training data, especially when the memory length is long. Large training overhead will decrease the transmission efficiency in the case of burst frame, multi-user links, or dynamic channel conditions [76, 77], where the taps need to be updated within several μ s or ms intervals.

To further improve the VNLE performance and reduce the requirement on training data size, in this dissertation, we investigate both the semi-supervised and supervised VNLEs with Lasso and perform an experimental validation in a fiber-FSO converged transport system [78, 79]. We leverage Lasso and ordinary least square (OLS) in the VNLEs and propose to use either the hard decision (HD) or soft decision (SD) in the semi-supervised VNLEs. Note that one benefit of Lasso is to reduce tap coefficients, which has been comprehensively investigated in supervised VNLEs and can help to reduce the implementation complexity [72, 73]. Another benefit of Lasso is that it can reduce the training symbol size requirement, especially when the ambient dimension of data vectors is much larger than the number of observations, which relaxes the latency and reduces training complexity. The supervised VNLE utilizes only the pilot symbols (labeled data). The semi-supervised VNLE utilizes both the labeled data and part of the unlabeled data to do polynomial regression based on pseudo-label [80, 81].

1.4.5 Ultra-reliable mmW and FSO system design

mmW and FSO are sensitive to different weather and atmospheric conditions. For instance, FSO is sensitive to fog and haze, while mmW suffers from rain, interference from other mmW channels, and frequency selective channel [18, 19, 82]. While the system can operate with feedback channel such as switching based schemes, it yields high complexity[83]. Adaptively combining the two complementary links, mmW and FSO, can enhance the reliability of the hybrid wireless system under different weather conditions without feedback complexity, as was previously reported [20]. However, the adaptive-diversity-combining technique (ADCT)-based system previously developed fails to address the needed tolerance to power attenuation in FSO link and frequency/time-burst distortion in mmW link. Moreover, the previous integrated mmW-FSO system underutilized the FSO bandwidth to fit with the narrower bandwidth of mmW, leading to a significant waste of bandwidth resources.

In this dissertation, we propose and experimentally demonstrate a coordinated mapping and combining (CMC) technique that enables the implementation of an ultra-reliable integrated mmW/FSO Analog RoF system [84]. The CMC technique exploits the abundant FSO and mmW bandwidth and shows superior performance in reliability and sensitivity compared to the ADCT-based system.

CHAPTER 2. FLEXIBLE SHAPING UNDER DIVERSE CHANNEL CONDITIONS

CHANNEL CONDITIONS

In this chapter, we will present the adaptive shaping techniques under diverse channel conditions, which result in more flexible systems and better network resource usage. We will experimentally investigate the shaping techniques in multi-carrier systems, fiber-FSO link based on IM-DD scheme, and next-generation flexible rate PON [38-42].

2.1 Definitions of Basic System Parameters

In this section, by taking PAS scheme as an example, we review the definitions of spectral efficiency, information rate, and generalized mutual information (GMI) for performance evaluation.

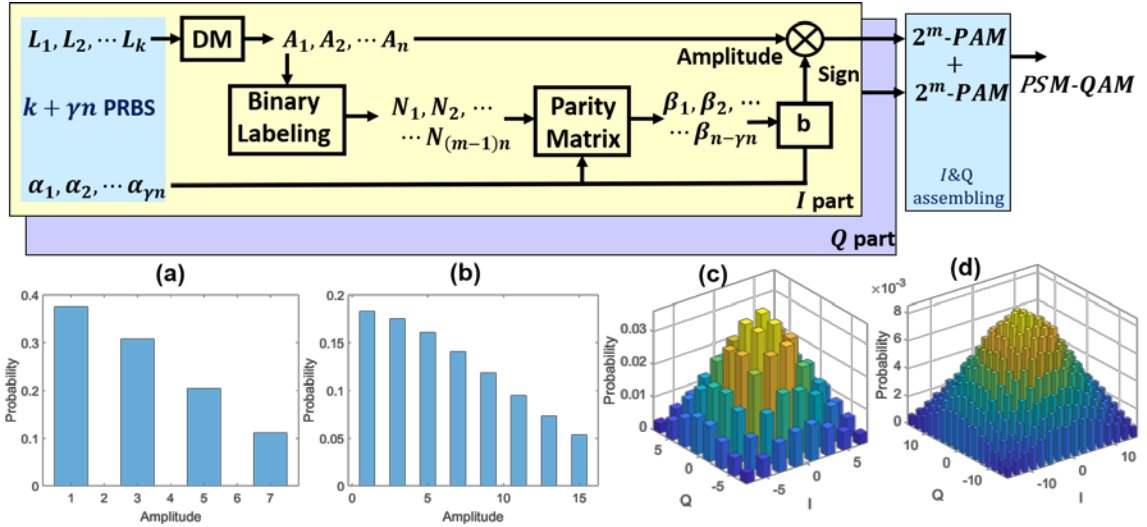


Figure 2. 1: Block diagram of the PAS scheme. Probability distribution of the amplitude part of (a) PS 64QAM, and (b) PS 256QAM. 2-D probability distribution of the assembled symbols of (c) PS 64QAM, and (d) PS 256QAM. © IEEE 2020

PAS is the state-of-art scheme to preserve the target distribution from the DM with FEC encoding for bi-polar distributions as discussed in the previous chapter. PAS works

for the case that the 2-D distribution is symmetric along the I axis and Q axis, and the probabilities of I and Q are independent with each other. Figure 2. 1 illustrates the block diagram of the PAS scheme. M- quadrature amplitude modulation (QAM) symbols require two parallel PAS blocks to generate I part and Q part (two PAM signal), respectively. Firstly, a pseudorandom binary sequence (PRBS) with a length of $k + \gamma n$ is divided into sequence 1 with a length of k and sequence 2 with a length of γn , respectively. Sequence 1 is loaded into the DM to generate the desired amplitude distribution. Given the targeted probabilities of the I part ($P_I(A)$) and Q part ($P_Q(A)$) in a PS M-QAM symbol, a target entropy of the QAM constellation can be expressed as $\mathbb{H}(X) = \mathbb{H}(P_I) + \mathbb{H}(P_Q)$, where X is the transmitted signal. For each PAS block in Figure 2. 1, the target entropy of the DM is $\mathbb{H}(P_I) - 1$ or $\mathbb{H}(P_Q) - 1$ with one bit serving as sign part. Most experimental demonstrations employ MB distribution to achieve enhanced Euclidean distance in an averaged power limited system. The probabilities are proportional to $\exp(-\lambda \cdot A^2)$, where A denotes the symbol amplitude. Thus, in this special case, I and Q follow the same distribution and are independent, which yields $\mathbb{H}(P_I) = \mathbb{H}(P_Q) = \frac{1}{2} \mathbb{H}(X)$. First, the output of DM serves as the amplitude of the PS-PAM symbol. In particular, the rate of the output of the CCDM is [29]

$$R_{dm} = (n_1 + n_3 + \dots + n_{\sqrt{M}-1})^{-1} \log_2 \frac{(n_1 + n_3 + \dots + n_{\sqrt{M}-1})!}{n_1! n_3! \dots n_{\sqrt{M}-1}!}, \quad (1)$$

where n_i is the number of symbols with each amplitude level in a PAS block. This DM rate is known to be strictly smaller than the target entropy $\mathbb{H}(P_{I,Q}) - 1$. On the other hand, the unconstrained bits that are mapped to the signs of the symbols induce no rate losses in the

PAS scheme. Putting them together and assembling I, Q part, we know that the number of bits per symbol is $2(R_{dm} + 1)$, which normalizing with $2m$ leads to the shaping set rate

$$R_{ss} = \frac{R_{dm} + 1}{m}. \text{ Consequently, the total rate loss (defined as bits/symbol) of the QAM}$$

symbols induced by CCDDM is:

$$\Delta_{PS} = 2\mathbb{H}(P_I) - 2(R_{dm} + 1) = \mathbb{H}(X) - 2(R_{dm} + 1). \quad (2)$$

When applying an FEC scheme, the overhead from parity bits is induced. Let R_{fec} denote the FEC code rate. As shown in Figure 2. 1, for the real parts of symbol sequence of length n , the amplitudes are mapped to a sequence of $(m-1)n$ bits to apply the FEC scheme, and γn bits are used to encode actual data in the signs. A subsequence of length $(1-\gamma)n$ in the signs are used for FEC redundancy. By symmetry, the same relation holds for the imaginary part. Therefore, the FEC code rate satisfies the following identity:

$$R_{fec} = \frac{1}{1 + OH_{fec}} = \frac{m-1+\gamma}{m}. \quad (3)$$

Note that since the FEC code rate minimum threshold is $\frac{(m-1)}{m}$ [85], the FEC overhead is

no more than $\frac{1}{m-1}$. (e.g., FEC overhead of PS 256QAM ($m=4$) should be $<33\%$). The

spectral efficiency of the entire system can then be written as:

$$SE = 2(R_{dm} + \gamma) = (R_{ss} - (1 - R_{fec})) \cdot \log_2 M. \quad (4)$$

M is the constellation order (e.g., $M = 4$ for PAM4, $M = 16$ for 16QAM). The cost from the FEC component is subtracted from the PS rate in the final expression of spectral efficiency (SE). To compare with the target signal entropy, the spectral efficiency can be further written as [86]:

$$SE = \mathbb{H}(X) - \Delta_{PS} - (1 - R_{fec}) \log_2 M. \quad (5)$$

where the signal entropy is $\mathbb{H}(x) = -\sum_i P(x_i) \cdot \log_2 P(x_i)$. In above expression, two terms are separately subtracted from the entropy, standing for CCDM rate loss and FEC cost, respectively. With a sufficient symbol length, the rate loss from DM can be very small [29]. The second part is $(1 - R_{fec}) \cdot \log_2 M$, which corresponds to the redundancy from FEC encoding. The information rate (IR) is defined as [38]:

$$IR = SE \cdot v, \quad (6)$$

where v is signal baud rate.

On the other hand, some research works have shown that GMI and normalized generalized mutual information (NGMI) [87] can be used for post-FEC performance evaluation with soft decision. Let the received data be Y , the achievable rate can be characterized as GMI [88]:

$$R_{BMD} = \max(\mathbb{H}(X) - \sum \mathbb{H}(B_i | Y), 0), \quad (7)$$

where B_i represents the i^{th} bit of the symbol X . GMI denotes the maximum number of information bits per transmit symbol in the bit-interleaved coded modulation additive white

Gaussian noise (AWGN) auxiliary channel, with ideal binary FEC decoding. NGMI implies the maximally usable FEC code rate with ideal binary FEC decoding and is expressed as [87]:

$$NGMI = 1 - (\mathbb{H}(X) - R_{BMD}) / m, \quad (8)$$

where m denotes bits per symbol (e.g., 2 for PAM4, 3 for PAM6/8). A practical FEC code has a gap Δ between its code rate R_{fec} and the NGMI, which is expressed as:

$$R_{fec} = NGMI - \Delta. \quad (\Delta \geq 0). \quad (9)$$

In the following analysis, we will use these metrics for performance evaluation of PS signal.

2.2 Entropy Allocation in Fiber-mmW Link

In the mmW RoF fronthaul, the orthogonal frequency-division multiplexing (OFDM) signal suffers from a non-uniform channel response due to frequency selective power fading or local oscillator leakage [89]. Therefore, increasing the link reliability and dealing with the uneven channel frequency response are critical for fiber-mmW converged links. It is common to track the SNR fluctuations over frequency based on bit loading, where different constellation orders are applied to different subcarriers [90]. Nevertheless, such a technique only changes the entropy by integers. Some reported works apply a uniform entropy (bits/symbol) to the whole OFDM frame with a performance improvement as shown in Figure 2. 2(a) [34, 91]. However, part of the subcarriers in an OFDM signal may be degraded or distorted significantly due to an uneven frequency response channel. To solve the issue, PS provides a continuous tuning of subcarrier entropy to an arbitrary, non-integer value using MB distribution [92-94]. The scheme has been applied to each

subcarrier to achieve maximum capacity. However, due to the per-subcarrier tuning scheme, it is prone to increase the network complexity and induce additional latency. Also, in the case of fixed interfaces such as fixed-rate transceiver, we need to allocate the entropy under a fixed total spectral efficiency. Therefore, we propose to divide the OFDM subcarriers into several PS units as shown in Figure 2. 2(b), and then assign different entropies to PS units based on their SNR conditions with a fixed total spectral efficiency constraint [38-40]. By grouping several subcarriers into one PS unit, we can relax the symbol length per subcarrier (which reduces latency requirement) and achieve flexible tuning of power margins while controlling algorithm complexity.

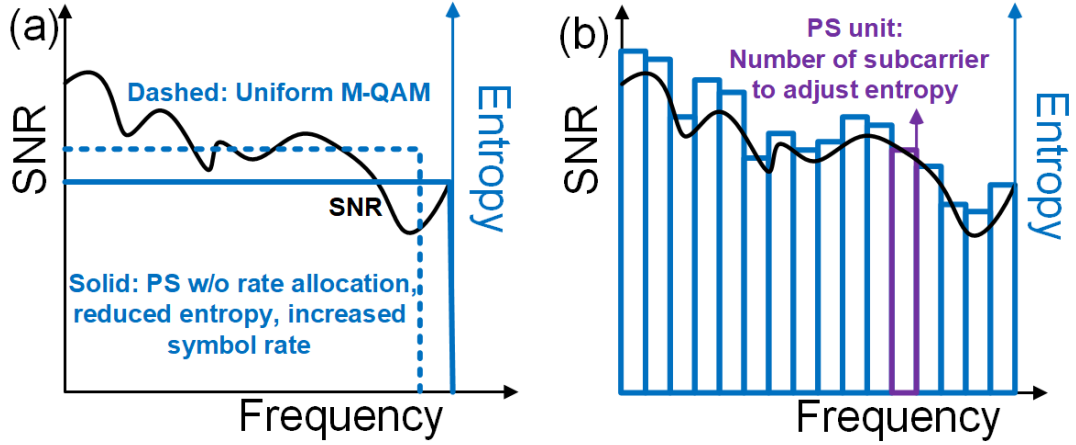


Figure 2. 2: PS unit-based entropy allocation. (a) Uniform M-QAM or PS with no entropy allocation. (b) Entropy allocation based on PS unit. ©IEEE 2020

2.2.1 Principle of Operation

This section presents the channel modelling of the PS QAM signal and the entropy (bits/symbol) optimization scheme. Note that in our analysis, a maximum likelihood (ML) detector, which makes the decision based on the minimum distance, is assumed. The modelling can be further extended to a maximum a posteriori (MAP) detector, which gives better performance but a higher complexity [95]. Firstly, an M-QAM signal is divided into

three parts: inner symbols, edge symbols and corner symbols. Figure 2. 3 presents an example of the division scheme for a 64-QAM signal.

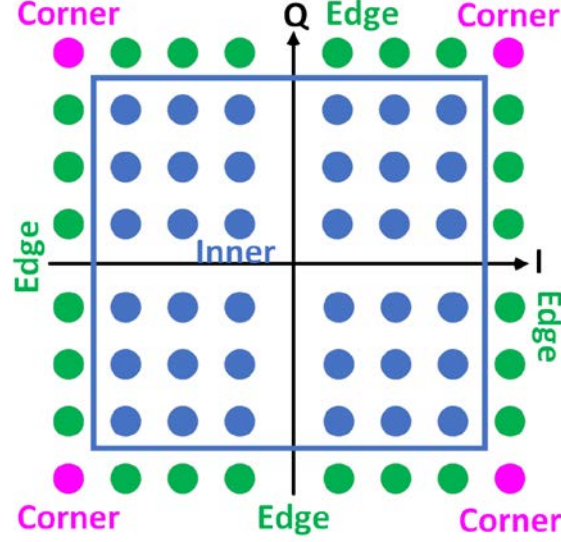


Figure 2. 3: The 64QAM constellation diagram of the inner, corner and edge symbols in modeling. ©IEEE 2020

Let the amplitude of the M-QAM signal be a_i and a_q (\sqrt{M} PAM alphabet) for I and Q part respectively, we can write the signal with AWGN as [96]:

$$r = \sqrt{E_g} \cdot \begin{bmatrix} a_i \\ a_q \end{bmatrix} + \begin{bmatrix} n_i \\ n_q \end{bmatrix} \quad (10)$$

where n_i and n_q are i.i.d. $N(0, N_0/2)$, and E_g is the energy of the received pulse. The statistics of the noise complex envelope shows white Gaussian noise with PSD equals to $N_0/2$. Then we can calculate symbol error rate (SER) as a function of E_g and N_0 .

Let the transmitted signal obey Maxwell-Boltzmann distribution in the PS scenario, the SER can be represented as the weighted sum of error probabilities of different parts:

$$\begin{aligned}
P_e &= 4 \cdot \frac{\exp\left(-2\lambda(\sqrt{M}-1)^2\right)}{Z} \cdot P(\text{wrong} \mid \text{corner}) \\
&+ \left(8 \cdot \sum_{i=1}^{\sqrt{M}/2-1} \left(\frac{\exp\left(-\lambda\left((\sqrt{M}-1)^2 + (2i-1)^2\right)\right)}{Z} \right) \right) \cdot P(\text{wrong} \mid \text{edge}) + \\
&\left(4 \cdot \sum_{j=1}^{\sqrt{M}/2-1} \sum_{i=1}^{\sqrt{M}/2-1} \left(\frac{\exp\left(-\lambda\left((2j-1)^2 + (2i-1)^2\right)\right)}{Z} \right) \right) \cdot P(\text{wrong} \mid \text{inner}) \\
&= \alpha_1 \cdot P(\text{wrong} \mid \text{corner}) + \alpha_2 \cdot P(\text{wrong} \mid \text{edge}) + \alpha_3 \cdot P(\text{wrong} \mid \text{inner}),
\end{aligned} \tag{11}$$

where the weight coefficients are functions of λ and correspond to the probabilities of each part. $P(\text{wrong} \mid \text{corner})$, $P(\text{wrong} \mid \text{edge})$ and $P(\text{wrong} \mid \text{inner})$ represent the error probabilities of corner symbols, edge symbols and inner symbols, which can be obtained as a function of $Q(\sqrt{2E_g/N_0})$ as shown below. The error probability of the corner symbols can be written as:

$$\begin{aligned}
P(\text{wrong} \mid \text{corner}) &= 1 - P(\text{correct} \mid \text{corner}) \\
&= 1 - P(n_I < \sqrt{E_g} \text{ \& } n_Q < \sqrt{E_g}) = 1 - P(n_I < \sqrt{E_g}) \cdot P(n_Q < \sqrt{E_g}) \\
&= 1 - \left(1 - Q(\sqrt{2E_g/N_0})\right)^2 = 2Q(\sqrt{2E_g/N_0}) - Q^2(\sqrt{2E_g/N_0}),
\end{aligned} \tag{12}$$

where Q is the Q function. Similarly, we can also obtain the error probability of the symbol for edge case and inner case, respectively [96].

$$\begin{aligned}
P(\text{wrong} \mid \text{edge}) &= 1 - P(|n_I| < \sqrt{E_g} \text{ \& } n_Q < \sqrt{E_g}) \\
&= 3Q(\sqrt{2E_g/N_0}) - 2Q^2(\sqrt{2E_g/N_0}).
\end{aligned} \tag{13}$$

$$\begin{aligned}
P(\text{wrong} \mid \text{inner}) &= 1 - P(|n_I| < \sqrt{E_g} \text{ \& } |n_Q| < \sqrt{E_g}) \\
&= 4Q(\sqrt{2E_g/N_0}) - 4Q^2(\sqrt{2E_g/N_0}).
\end{aligned} \tag{14}$$

Z is the normalization term, which is $4 \cdot \sum_{i=1}^{\sqrt{M}/2} \sum_{j=1}^{\sqrt{M}/2} \left(\exp \left(-\lambda \left((2j-1)^2 + (2i-1)^2 \right) \right) \right)$. Moreover,

$Q(\sqrt{2E_g/N_0})$ can be represented with signal SNR and the mean value of signal power. The energy per bit can be written as

$$E_b = \frac{\text{Energy per pulse}}{\text{entropy}} = \frac{\mathbb{E}(a_i^2 + a_q^2)}{\mathbb{H}} E_g, \quad (15)$$

where \mathbb{H} is the signal entropy and the mean value is

$$\begin{aligned} & \mathbb{E}(a_i^2 + a_q^2) \\ &= \sum_{i=1}^{\sqrt{M}/2} \sum_{j=1}^{\sqrt{M}/2} \left(\frac{\exp \left(-\lambda \left((2j-1)^2 + (2i-1)^2 \right) \right)}{Z} \cdot \left((2j-1)^2 + (2i-1)^2 \right) \right). \end{aligned} \quad (16)$$

With $E_b/N_0 = \text{SNR}/\mathbb{H}$, the Q function can be written as:

$$\begin{aligned} Q(\sqrt{2E_g/N_0}) &= Q\left(\sqrt{2E_b \cdot \mathbb{H} / (N_0 \cdot \mathbb{E}(a_i^2 + a_q^2))}\right) \\ &= Q\left(\sqrt{\frac{2E_b}{N_0} \cdot \frac{\mathbb{H}}{\mathbb{E}(a_i^2 + a_q^2)}}\right) = Q\left(\sqrt{\frac{2\text{SNR}}{\mathbb{E}(a_i^2 + a_q^2)}}\right). \end{aligned} \quad (17)$$

Combining equation 11 with equation 17, we obtain P_e as a function of SNR, λ , and modulation order M . The bit error rate (BER) can be approximated as $P_e/\log_2 M$ with gray mapping [96]. Note that λ can also be represented with signal entropy:

$$\mathbb{H} = 4 \cdot \sum_{i=1}^{\sqrt{M}/2} \sum_{j=1}^{\sqrt{M}/2} \frac{\exp \left(-\lambda \left((2j-1)^2 + (2i-1)^2 \right) \right)}{Z}. \quad (18)$$

Finally, we can get Q factor as a function of SNR, signal entropy and QAM order M with

$$Q^2 (dB) = 20 \log_{10} \left(\sqrt{2} \operatorname{erfc}^{-1} (2BER) \right) [97].$$

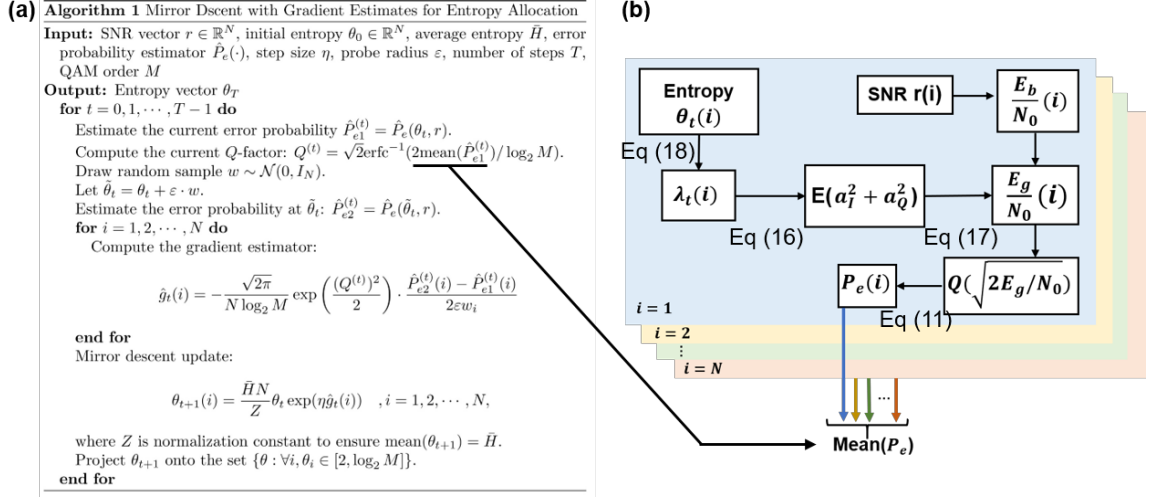


Figure 2. 4: Entropy allocation algorithm. (a) Modeling based algorithm for entropy allocation. (b) Chart flow for calculating P_e given SNR and entropy. ©IEEE 2020

Figure 2. 4 illustrates the entropy optimization algorithm based on the model that we built up in the previous analysis. Most of the commercial wireless systems operate with multicarrier waveform, i.e., OFDM, in which each subcarrier occupies small bandwidth. Therefore, each subcarrier can be approximated by AWGN channel due to smaller bandwidth even if the received signal suffers from non-uniform frequency responses. On the other hand, the operating ROP point of practical system (nearby the pre-FEC thresholds) is dominated by ASE noise. This implies that the model built with AWGN assumption can still work. Assume that there are N PS units (each PS unit is comprised of few subcarriers), we can obtain the pre-measured channel SNR vector $r \in \mathbb{R}^N$ for every PS unit. One constraint of the optimization algorithm is that average entropy of the whole signal frame equals target entropy \bar{H} . Given r, \bar{H} , the initial entropy vector θ_0 , step size η , probe

radius ε and number of steps T , we can compute the allocated entropy step by step, as shown in Figure 2. 4.

We use a projected mirror descent algorithm for maximizing the Q factor (thus the averaged SER) among the PS units, where the gradient information is estimated by taking finite difference for function value information using the modeling result. Note that we need to estimate the gradient of the Q-factor with respect to entropy of each PS unit. This can be implemented via chain rule: we can compute $\partial Q / \partial P_e(i)$ analytically, and compute $\partial P_e(i) / \partial \theta_i$ numerically by finite difference. Multiplying them together, we obtain the gradient estimator \hat{g} in the Algorithm described by Figure 2. 4(a). Mirror descent is a variant of gradient descent which adapts to the geometry of the problem and enforces the constraint conveniently, with possibly faster convergence rate [98]. In general, for any $x, y \in R^n$, the algorithm uses a convex mirror map function Φ to define the Bregman divergence $D_\Phi(y, x) = \Phi(y) - \Phi(x) - \langle y - x, \nabla \Phi(x) \rangle$. Suppose we are given a step size η and a gradient estimator g . Let x^t be the current iterate and x^{t+1} be the next iterate, the algorithm takes the update:

$$x^{t+1} = \arg \min_y \eta \langle g(x^t), y - x^t \rangle + D_\Phi(y, x^t), \quad (19)$$

where $\langle g(x^t), y - x^t \rangle$ denotes the inner product between the vectors $g(x^t)$ and $y - x^t$. In order to enforce the constraint that the average entropy \bar{H} over channels is fixed, we choose

the mirror map function to be $\Phi(x) = \sum_{i=1}^n \frac{x_i}{N\bar{H}} \log \frac{x_i}{N\bar{H}}$. In this case, the function $D_\Phi(x, y)$

is the Kullback-Leibler (KL) divergence between $x / N\bar{H}$ and $y / N\bar{H}$, the normalized

entropy vectors seen as probability distribution over N coordinates. The one-step update defined above can be written in closed form as exponential weight update, which is shown in Algorithm 1.

Define the norm $\|x\|_1 = \sum_i |x_i|$. It is known that within a domain of diameter R under the $\|\cdot\|_1$ metric, for a convex functions f such that: $\max_i |\partial f / \partial x_i(x)| \leq L$, the minimization of f converges to the optimal solution with rate RL / \sqrt{T} [98]. Though the objective function in the entropy allocation problem is generally non-convex, it is locally convex in a region around a local minimum. Thus, an initialization point acquired by the heuristics-based allocation method satisfies the convergence rate in the local region. For a system with N PS units, the diameter R under the $\|\cdot\|_1$ metric scales linearly with N , as entropies are bounded. Note that in our model, the one-dimensional mapping from SER to the Q factor and the mapping from entropy to SER both have bounded derivatives in a bounded region, the Lipschitz constant L scales with $1/N$, due to the averaging of SER per PS unit. Overall, the algorithm has a rate of convergence independent of number of PS units. This is made possible by the mirror descent algorithm which exploits the structure of the objective function and the constraint set. Finally, the mirror descent algorithm directly enforces the average entropy constraint that the averaged entropy of N PS units equals to \bar{H} . In each iteration of Algorithm 1, we perform a mirror descent update. To estimate the gradient of Q, we sample an entropy vector near the current vector and calculate its SERs per channel and final Q factor. We can take the difference between the old and new SERs and Q factors, from which the gradient of Q can be computed. Note that given an entropy vector, we can compute the signal SER based on the block diagram in Figure 2. 4(b). The

current signal error probability can be obtained from taking average of the SER of the N PS units.

2.2.2 Experimental Setup

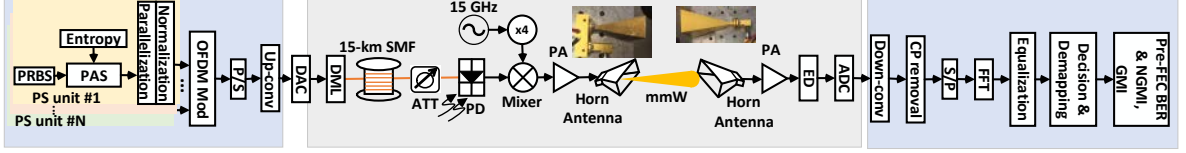


Figure 2. 5: Experimental setup and DSP diagram of the fiber-mmW system with entropy allocation. ©IEEE 2020

Figure 2. 5 illustrates the experimental setup and DSP diagram. In the transmitter DSP, subcarriers in the same PS unit share the same entropy and they are combined together to generate the target entropy with a smaller rate loss. For every PS unit, a PRBS is loaded to a PAS block with the assigned entropy from the optimization algorithm. In the PAS block, the CCDM generates the desired probability distribution for the amplitude part of the I and Q signal, respectively. We choose the target distribution as the Maxwell-Boltzmann distribution. Then the amplitude part combines with the sign part, which gives the QAM signal. The symbol sequence in each PS unit is then parallelized into several subcarriers and normalized. In this demonstration, the PS-64QAM OFDM signal consists of 800 subcarriers and 54 symbols with a subcarrier spacing of 750 kHz, while PS-256QAM OFDM has a similar setting but with a subcarrier spacing of 500 kHz. The Inverse fast Fourier transform (IFFT) size is 2048 for both schemes and 5.12% Cyclic Prefix (CP) is inserted. After OFDM modulation (IFFT operation and CP insertion) and P/S conversion, the complex OFDM signal is up converted to an intermediate frequency (IF) signal with a carrier frequency of 500 MHz. The IF signal is loaded into a DAC with

a sampling rate of 16 GS/s and 8-bits effective number of bits (ENOB). Then it is modulated by a 10-GHz directly modulated laser (DML). The optical signal is delivered through a 15-km single-mode fiber (SMF) and detected by a 10 GHz PD. An attenuator before the PD controls the ROP in the system. The electrical IF signal is then converted to a 60-GHz mmW signal by a 18-GHz electrical mixer (input third order intercept equals to 13 dBm), a 15-GHz RF source and a frequency quadrupler. A wideband power amplifier (PA) are employed to boost the signal on both sides of antennae. A horn antenna (50GHz to 75GHz) with 25-dBi gain radiates the 60-GHz radio signal over the air while a symmetric horn antenna receives the signal after 1.5-m wireless transmission. A 1-GHz envelope detector (ED) down-converts the radio signal to the IF signal again. Then a 5-GS/s ADC recovers the digital waveform. In the receiver DSP, we first down-converts the IF signal to baseband signal. After CP removal, S/P conversion, FFT and equalization with a one-tap equalizer, we perform the decision and de-mapping, which can be utilized to calculate the pre-FEC metrics such as BER, GMI and NGMI.

Table 2. 1: SE of different signals with different OHs. ©IEEE 2020

	7% FEC OH	16.3% FEC OH
SE of uniform 16QAM	3.74 bits/symbol	3.44 bits/symbol
SE of PS 4.28 64QAM	3.89 bits/symbol	3.44 bits/symbol
SE of uniform 64QAM	5.61 bits/symbol	5.16 bits/symbol
SE of PS 6.28 256QAM	5.76 bits/symbol	5.16 bits/symbol

We utilize two pre-FEC thresholds in the analysis: 1) pre-FEC BER equals to (HD-FEC) with 7% FEC overhead (OH) (code rate = 0.935 [99]) NGMI threshold of 0.86 (ideal FEC) with 16.3% FEC OH (code rate = 0.86[100]). Considering the rate loss is negligible,

we choose PS 64-QAM signal (4.28 bits/symbol) versus uniform 16QAM, and PS 256 QAM (6.28 bits/symbol) versus uniform 64QAM. The SEs with the two FEC OH are summarized in Table 2. 1. Note that with 3.8×10^{-3} pre-FEC threshold (OH 7%), the SE of the PS signal is slightly higher than the uniform signal. In the experiment, the DM rate losses of the PS 64QAM and PS 256QAM are 0.038 bits/symbol and 0.098 bits/symbol, respectively. However, in our following analysis, the PS signal with 7% OH (with higher SE) still shows sensitivity gain compared with uniform signal.

2.2.3 Computation Complexity and Parameter Setting

In the following, we analyze the computational complexity of the proposed algorithm. Firstly, the number of iterations is independent of the number of PS units as illustrated previously. Therefore, we only need to study the per-iteration cost for the algorithm. In each iteration, we compute the error probability for each PS unit independently and calculate the averaged Q factor twice, by taking the difference of which we estimate the gradient. The computational complexity scales linearly with N , the number of PS units in the system. For each PS unit, we need to carry on the computation in equation 11-17 for a given entropy level, which takes the computational complexity scaling linearly with the QAM order M . Overall, the computational complexity for the algorithm is $O(MN)$. It is worth noting that the computations in each channel and the computation related to each (i, j) pair can be implemented in parallel, making it potentially fast with a hardware implementation.

In the experiment, we set η as 0.1 and ε as 10^{-9} . Firstly, we investigate the number of steps (iteration) T based on the pre-measured SNR in the setup. Figure 2. 6 demonstrates

root mean square (RMS) of the difference between the current allocated entropy vector θ_{t+1} and the allocated entropy vector θ_t in previous iteration with the increasing of the iteration number T . The predicted Q factor in every iteration is also presented, which aligns with the trend of the RMS of entropy vector difference. One can note that the variance of the Q factor is relatively small after 15 times of iterations while the allocated entropy RMS converges when T is around 20. In the experiment, T is fixed as 30 to ensure that the convergence results are always achieved.

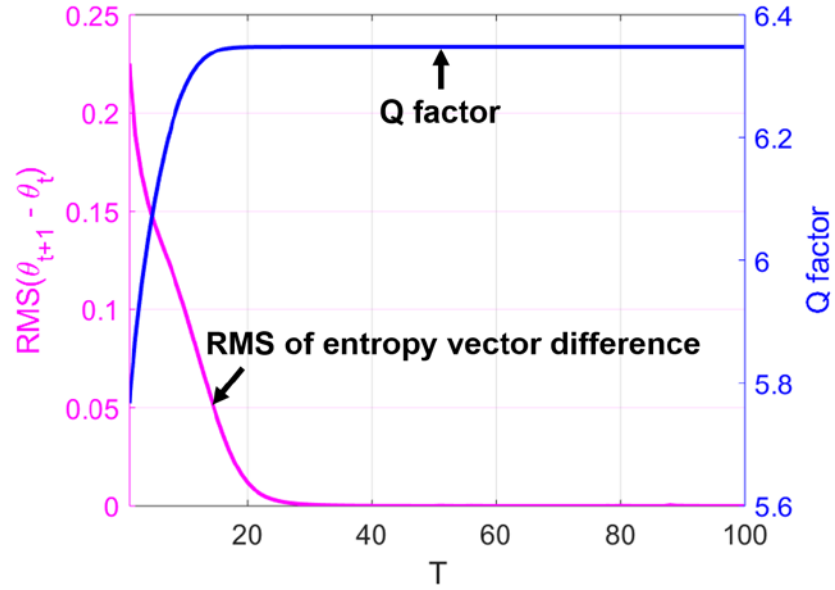


Figure 2. 6: Convergence speed of the entropy allocation algorithm. ©IEEE 2020

Note that one subcarrier loading might give better entropy loading resolution but yield significant rate loss in our experiment setup. Thus, to make a trade-off between entropy loading resolution and rate loss optimization, PS unit size is fixed as 10 subcarriers per unit in the following algorithm comparisons.

2.2.4 Experimental Results

The channel SNR distribution over frequency is pre-measured through the EVM of training signal via [101]:

$$SNR = 1/EVM^2. \quad (20)$$

The training symbols are 20 QPSK symbols. Based on the pre-measured SNR, we can compute the assigned entropy using different algorithms. Figure 2. 7 shows the allocated entropy result. The black curve is the pre-estimated SNR curve of the channel. The proposed algorithm can track the SNR fluctuations over the frequency and show similar shape to the SNR curve.

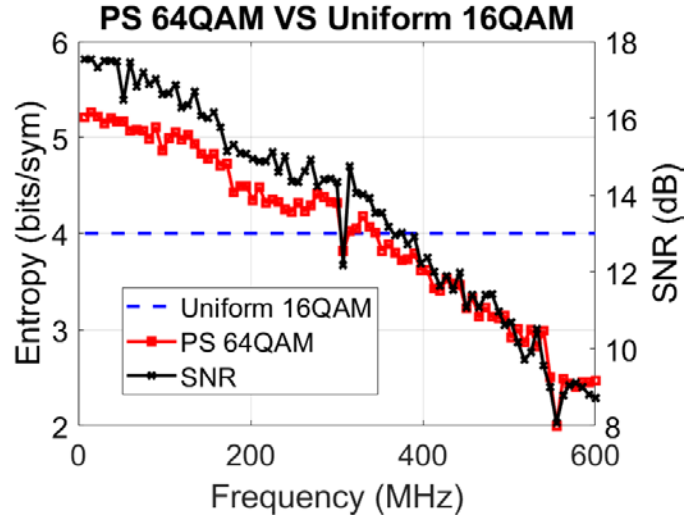


Figure 2. 7: SNR or Assigned entropy over frequency. ©IEEE 2020

Then we sweep the ROP and compare the sensitivity performance of the algorithms. Figure 2. 8(a) to (c) present the performance of PS 4.28 64QAM versus uniform 16QAM. In the experiment, the pre-FEC BER is calculated using ML hard decision. NGMI utilizes the bit-wise log-likelihood ratio and assumes an ideal binary FEC. In practice, the pre-FEC characterization parameter depends on the specific FEC scheme implemented in the system.

For instance, pre-FEC BER is utilized for HD-FEC while GMI is more suitable for SD-FEC.

Figure 2. 8 (a) shows that, compared with uniform 16QAM, the proposed algorithm shows a 1.8-dB gain with the BER threshold of 3.8×10^{-3} . Note that the SE or information rate of the PS signal is higher than that of the uniform signal in this case. The sensitivity gain could be higher if keeping the SE. Figure 2. 8(b) illustrates the GMI versus ROP. In comparison to the uniform 16QAM, our proposed algorithm shows a 0.4-bits/symbol improvement of GMI value when the ROP ranges from -4 dBm to -2 dBm. When the ROP is higher, the curves gradually converge due to saturation to its maximum supported entropy. The uniform 16QAM is always worse than the curves with entropy allocation. Figure 2. 8(c) compares the NGMI performances over frequency for three algorithms at the ROP of -3 dBm. Reference line of NGMI=0.86 is provided. PS-OFDM 64-QAM signal with the proposed algorithm has a flat NGMI curve. The schemes with uniform 16QAM cannot pass the 0.86 threshold in some frequency regions. For the sharp “dip” of the uniform 16QAM signal (where the signal has a very significant SNR degradation such as ~300 MHz or 550 MHz), the proposed algorithm still has a good NGMI.

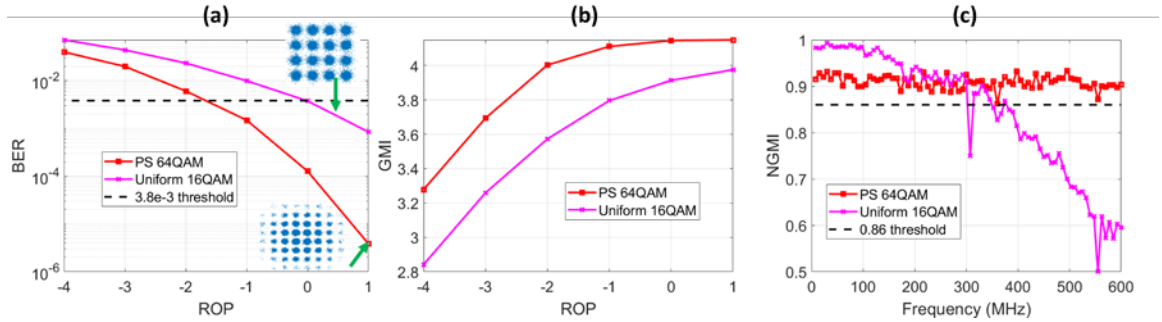


Figure 2. 8: Experimental results of PS 64QAM. ©IEEE 2020

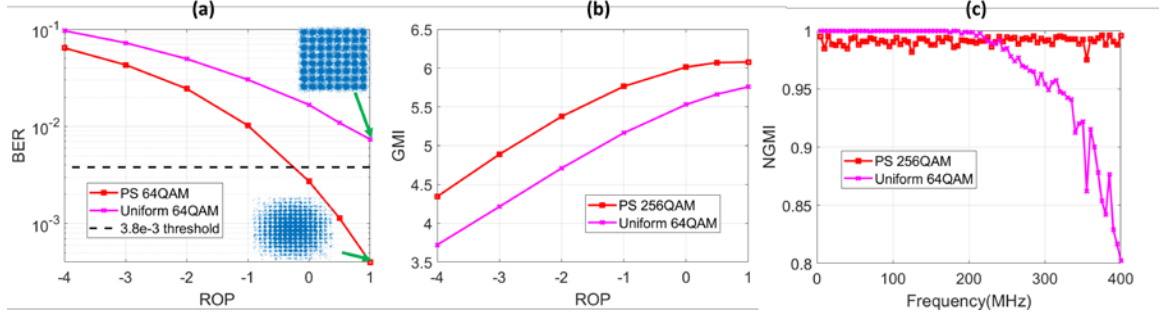


Figure 2. 9: Experimental results of PS 256QAM. ©IEEE 2020

As for PS 256QAM versus uniform 64QAM, the proposed algorithm has > 1.5 -dB sensitivity gain with respect to 3.8×10^{-3} pre-FEC BER threshold (Figure 2. 9(a)). As for GMI in Figure 2. 9(b), the proposed algorithm also has a 0.7 bits/symbol GMI improvement. As illustrated in Figure 2. 9(c), the channel response is also smoother using the PS 256QAM.

2.2.5 Summary

We have proposed a novel algorithm for entropy allocation in the PS-OFDM mobile fronthaul link. The closed-form expression of the pre-FEC SER versus SNR under AWGN channel condition is derived. Based on the modeling result, the optimization algorithm utilizes the pre-measured SNR information and allocates the entropy intelligently using a projected mirror descent algorithm with the pre-measured SNR information. Experimental demonstration indicates a convergence speed of 20 to 30 iteration times. In summary, the proposed algorithm is promising to extend the link power budget or improve the signal quality in a fiber-mmW converged fronthaul, especially in an uneven channel condition.

2.3 Shaping in High-Speed Fiber-FSO Link

With ample bandwidth and flexibility of deployment, FSO is a promising technology to provide last-mile data delivery in the absence of or in addition to fiber links [17]. Most fiber-FSO link utilizes IM-DD scheme due to its cost efficiency. Therefore, as mentioned in Chapter 1, the optical power corresponds to the first moment of the signal in a single-carrier IM-DD system, which results in unipolar exponential distributions to maximize the entropy. The parity bits are inserted into the LSB to form a PW distribution with negligible performance degradation in certain signal-to-noise ratio ranges as illustrated in Figure 1.7. In this section, we experimentally demonstrate a low-density parity-check (LDPC) coded PAM-4/8 transmission using the unipolar, PW exponential distribution in a fiber-FSO link [41]. Moreover, we employ nonlinear pre-distortion to further mitigate the transmitter nonlinearity. The receiver side utilizes a simple linear equalizer. We investigate two comparison schemes in the experiment. Experimental results have attained variable power margins improvement (up to 2.1-dB) when compared with the uniform distribution.

2.3.1 Principle of Operation

Figure 1.7 (a) presents the conventional PAS scheme, which inserts the parity bits into MSB and forms a bipolar MB distribution. However, the entropy-maximizing distribution for the IM-DD system with average optical power limitation is unipolar exponential distribution [36, 37]:

$$P(x_j) = \exp(-\lambda \cdot x_j) / \sum_i \exp(-\lambda \cdot x_i), \quad (21)$$

where x_j is the j^{th} symbol from the alphabet $[x_i]$ and λ can tune the signal distribution and entropy. The probability decreases monotonically with the increase of signal level. Direct

combination with LDPC distorts the probability distribution from DM, considering the parity bits are uniformly distributed. Thus, we generate half of the alphabet using DM based on the exponential distribution in equation 21. The parity bits and some unconstraint bits are placed in the LSB, which yields a PW distribution where two adjacent symbols share the same probability [12]. Thus, the distribution from the DM is well preserved.

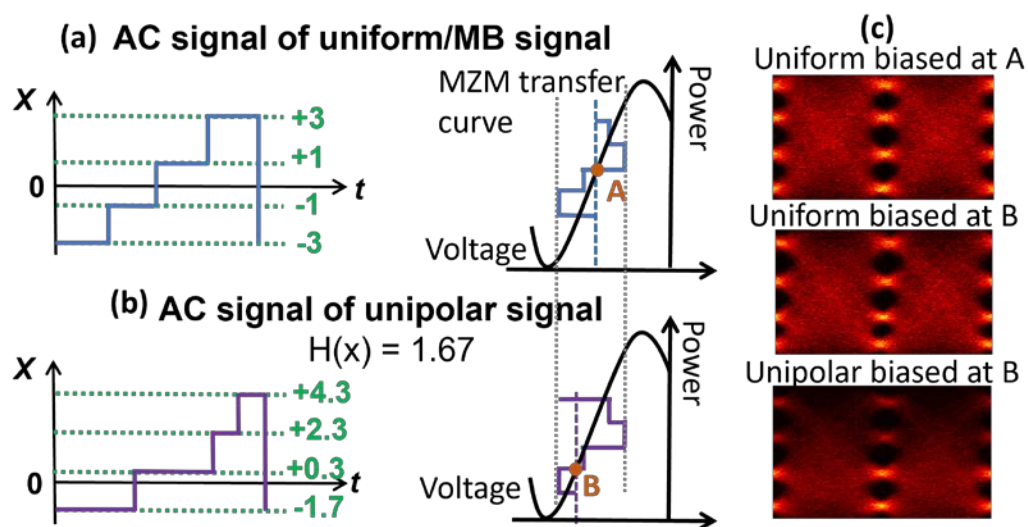


Figure 2. 10: Principle of operations of the PW PS PAM signal. ©OSA 2021

However, different from the uniform or MB signal, the unipolar alternating current (AC) signal is not symmetrical anymore as shown in Figure 2. 10. The direct current (DC) bias thus is closer to the lower signal level. The DC bias of unipolar signal (B) moves away from the quadrature point (A) and closer to the null point to maintain a similar extinction ratio (ER) to that of the bipolar signal as illustrated by Figure 2. 10(a) and (b). Figure 2. 10(c) presents the measured eye diagrams at different bias points. The uniform signal biased at point A yields a good eye diagram, but it becomes worse when biased at point B. However, the unipolar signal shows a good eye diagram when the bias point moves closer to the null point, to point B. Therefore, with the same ER and Euclidean distance, the

unipolar signal has lower average optical power. With the same optical power by either tuning the laser output power or using EDFA as a pre-amplifier, the Euclidean distance of the unipolar signal is larger than that of the bipolar signal.

In the following analysis, we investigate two comparison schemes in the experiment according to equation 5 and 6. One is tuning bandwidth (ν) to make sure the IR is the same with the same FEC code rate (R_{fec}), while the other is jointly adjusting the FEC code rate and signal entropy to make sure the SE is the same with a fixed bandwidth.

2.3.2 Experimental Setup

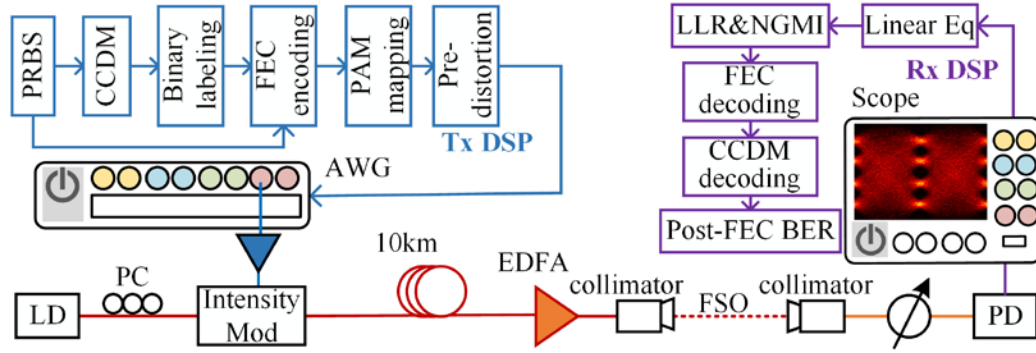


Figure 2. 11: Experimental setup of the fiber-FSO link based on unipolar distribution. ©OSA 2021

Figure 2. 11 presents the experimental setup. At the transmitter side, part of the bits sequence is loaded into the DM, while other bits form as the unconstrained bits. DM generates half of the symbol amplitude according to the exponential distribution. Then the DM output is binary labeled. Together with the unconstrained bits, the labeled DM output is loaded into the LDPC encoding block. The parity bits are in the LSB. The parity matrix is from the Digital Video Broadcasting standard DVB-S.2. Then the binary sequence is

mapped to PAM symbols. To simplify the receiver DSP, we add nonlinear pre-distortion at the transmitter side, which performs Volterra indirect learning [69]. The tap number for the linear term is 21. The polynomial terms are up to third order with memory length equals to 5. The training symbol size is 21600. After pre-distortion, the signal is loaded into Keysight 65-GSa/s arbitrary waveform generator (AWG), and then boosted by a 25-dB amplifier. The laser output power is 12 dBm, and the modulator bandwidth is 20 GHz. After intensity modulation, the signal is delivered through 10-km SSMF and amplified by an EDFA operated at the constant output power mode. Finally, the amplified optical signal passes through a 2.5-m FSO link with 5-dB loss and is detected by a PD. A 25-GHz, 80-Gs/s Keysight real-time scope captures the detected signal. In the receiver DSP, first, a linear equalizer recovers the received signal using 5% pilot symbols. Then we perform a soft decision to calculate the loglikelihood ratio (LLR) and NGMI [87], which serves as a pre-FEC performance metric. LLRs are loaded into the LDPC decoding block followed by the DM decoding and post FEC BER counting.

2.3.3 *Experimental Results*

We sweep and choose the peak-to-peak voltage from the AWG using the uniformly distributed signal in Figure 2. 12. Then we fix peak-to-peak voltage as 0.25 V for both PS signals and uniform signals. Figure 2. 13(a) and (b) compare the NGMI performances with the same LDPC code rate equals 0.8. PS signals with lower entropy have a higher baud rate, thus the same information rate for a fair comparison. NGMI indicates the required FEC code rate assuming ideal FEC. In practice, there is a gap between NGMI and FEC code rate, which can be obtained through Monte-Carlo simulation [87]. In Figure 2. 13(a), at the NGMI thresholds from 0.8 to 0.9, the unipolar PAM4 1.73 signal demonstrates a 1.1-

dB sensitivity gain. However, MB signals become worse than uniform PAM4 with the same peak-to-peak voltage due to the increased peak to average power ratio (PAPR). As for PAM8 in Figure 2. 13(b), the unipolar signals show the best performance, and the MB signals are slightly better than the uniform signals. Figure 2. 13(ii) shows that the lowest level and highest level of the uniform PAM8 signal suffer from nonlinearities. MB signal assigns a smaller probability to these two levels so that its eye diagram shows less nonlinearity (Figure 2. 13(iii)). The unipolar PAM8 2.4 demonstrates 2.1-dB ROP improvement when the NGMI ranges from 0.9 to 0.8.

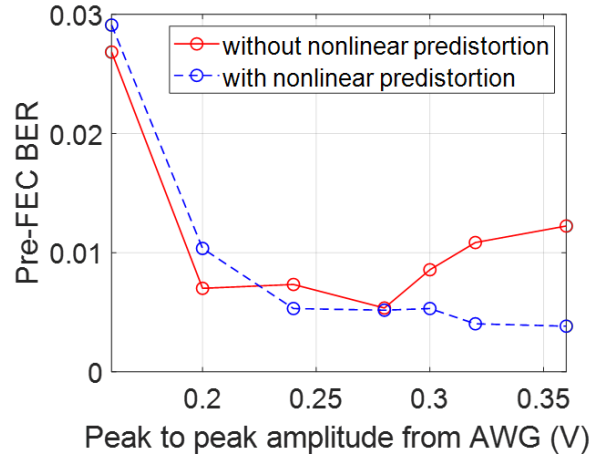


Figure 2. 12: Sweep peak-to-peak amplitude. © OSA 2021

Besides fixing the FEC code rate and tuning the baud rate, we can also adjust the signal entropy and FEC code rate to maintain the same SE with the same baud rate. In Figure 2. 14(a), the uniform PAM4 utilizes $R_{fec} = [5/6, 3/4]$, which yields two SE values. The unipolar signals use $R_{fec} = 8/9$ and can tune its entropy to match the two SE values of the uniform PAM4. At the post-FEC BER threshold of 1×10^{-4} , the unipolar signal demonstrates 0.7-dB and 0.4-dB sensitivity improvement at SEs of 1.67 and 1.5 bits/symbol, respectively.

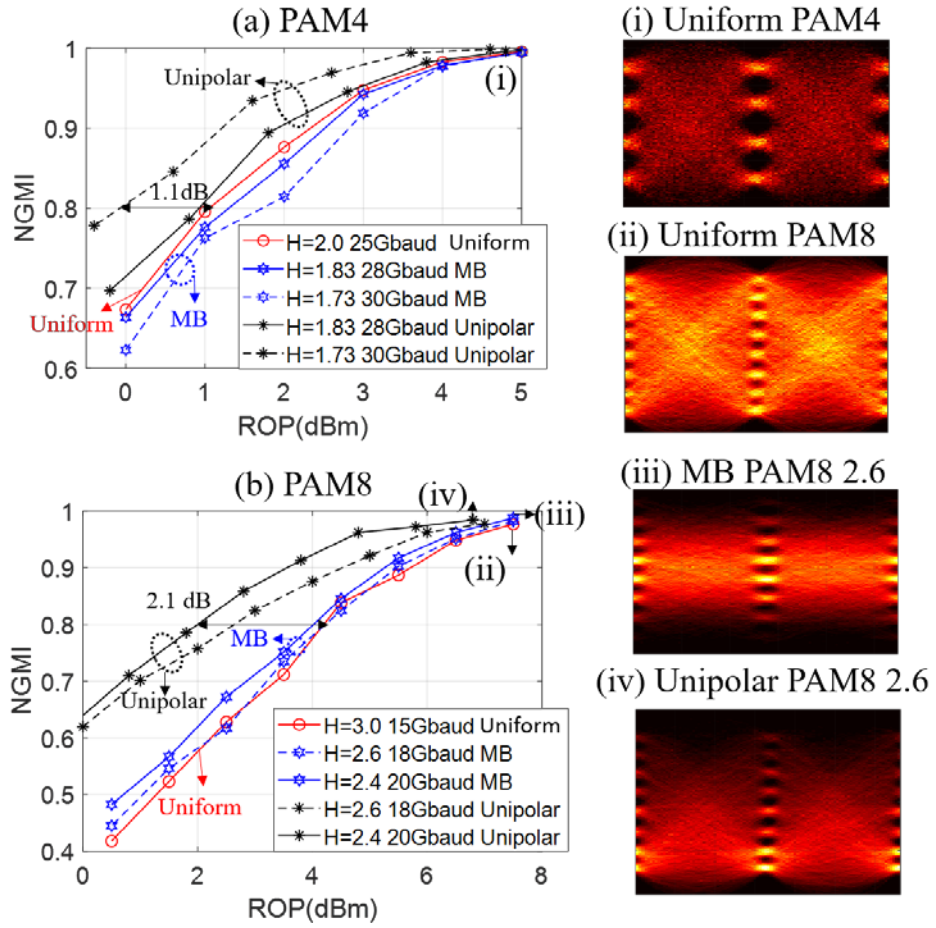


Figure 2. 13: Results with fixed FEC code rate and the same information rate. © OSA 2021

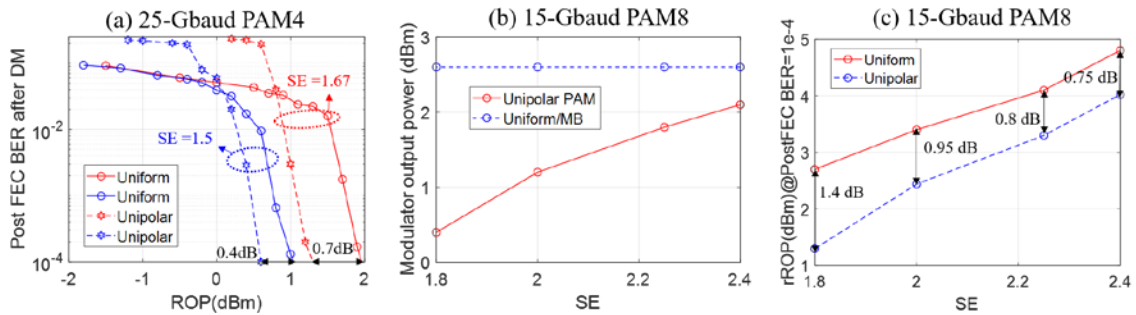


Figure 2. 14: Results with the same spectral efficiency and flexible FEC code rate. ©OSA 2021

As for uniform PAM8 in Figure 2. 14(b) and (c), we choose $R_{fec} = [4/5, 3/4, 2/3, 3/5]$, which gives four SE values. Unipolar PAM8 utilizes $R_{fec} = 5/6$ and adjusts its entropy to maintain the same SE as the uniform PAM8. The modulator power of the unipolar signal decreases with the decrease of SE, while that of the uniform signal remains the same. Moreover, at the post-FEC BER of 1×10^{-4} , the required ROP gain increases with the decreasing of SE. Compared with uniform PAM8, unipolar PAM8 achieves 0.75 to 1.4-dB improvement with SE varying from 2.4 bits/symbol to 1.8 bits/symbol.

2.3.4 Summary

In a summary, we experimentally demonstrated LDPC coded PAM transmission with unipolar distribution and pre-distortion in a fiber-FSO hybrid link with the parity bits inserted in the LSB. Unipolar distribution exhibits up to 2.1-dB sensitivity improvement in terms of the case of fixed FEC code rate and IR with adjustable signal baud rate. In the case of the same SE by tuning the FEC code rate and signal entropy, the unipolar signal demonstrates up to 1.4-dB ROP gain. In the next section, we will extend to a more complicated IM-DD scenario.

2.4 Shaping for Flexible Rate PON

Up till now PON standards have employed a fixed data rate designed for the worse-case channel conditions [4]. This leads to sub-optimal resource utilization. To allow for more flexibility to support the disaggregation of network functions in the optical transport layer and to better utilize resources, in this section, we investigate how to flexibly adapt the data rate of next generation PON systems depending on channel conditions of the users. A straightforward way to flexibly adjust the data rate is to change the modulation order

(e.g., NRZ, PAM4, and PAM8) for the different users [23]. However, entropy tuning using uniform-distributed signals is stepwise and improvement is hampered by the large sensitivity difference between different modulation formats. Therefore, we propose to adopt PS for continuous, flexible entropy tuning with enhanced noise tolerance. Some research papers have investigated IM-DD systems using unipolar exponential distribution by maximizing the entropy in data center applications [36]. However, since PON has specific requirements including large power budgets (min 29 dB) and using avalanche photodiodes (APDs), the interaction with noise (e.g., shot noise and thermal noise) and link budget optimization are still not well studied. Therefore, we therefore investigate PS specifically for PON in an IM-DD single-carrier system. To the best of our knowledge, this is the first study of applying PS and GS to next-generation PON systems.

2.4.1 Principle of Operation

We characterize the achievable rate as GMI in equation 7, which is denoted as R_{BMD} . In our case, GMI is calculated based on gray mapping. Let the transmitted data be X , α be the electrical to optical (E/O) coefficient of modulator, G be the semiconductor optical amplifier (SOA) gain, and P_{avg} be the average output power from SOA. The optimization problem is formulated as:

$$\begin{aligned} \max_{P_x, S_x} \quad & R_{BMD} \\ \text{subject to} \quad & \mathbb{E}[\alpha \cdot G \cdot X] = \alpha \cdot G \cdot \sum P_x \cdot S_x = P_{avg}, \\ & \max(\alpha \cdot G \cdot X) \leq X_{th}, \text{ and NGMI} \geq R_{th}, \end{aligned} \quad (22)$$

which is solved by the interior-point method. To avoid local maxima, we run the algorithm five times with random starting points. P_x and S_x denote probability distribution from PS

and signal levels from GS. X_{th} is SOA peak power threshold to maintain linear operation. R_{th} is the NGMI threshold considering practical implementation.

Figure 2. 15 illustrates the schemes investigated in this section using PAM4 as an example. Figure 2. 15(a) shows uniform PAM4 and (b) indicates a typical optimal probability distribution. The distribution is neither symmetric nor monotonic. The asymmetry makes it unsuitable for conventional serial combination with FEC or probabilistic amplitude shaping since FEC parity bits are typically uniformly distributed. Therefore, we use pairwise distribution as shown in Figure 2. 15(c) [12]. Parity bits placed in the LSB distinguishes the two levels in each pair with the same probability, thus preserving the DM distribution. Similar to PAS scheme, the pairwise constraint introduces a minimal code rate and thus minimal NGMI requirement as:

$$\text{NGMI} \geq R_{th} = (m-1) / m. \quad (23)$$

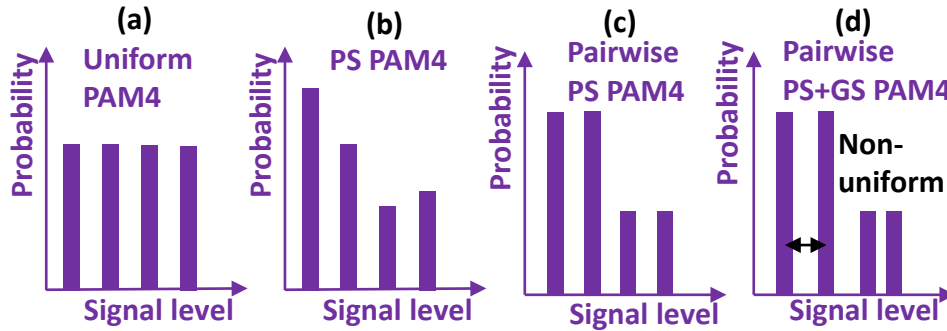


Figure 2. 15: Illustration of the investigated schemes using PAM4 as an example. © IEEE 2020

However, the pairwise constraint introduces a performance penalty compared to the optimal distribution. Therefore, we propose to use GS (Figure 2. 15 (d)) to mitigate the

penalty due to pairwise distribution. Moreover, we also investigate PS-PAM6, which is a reduced PS PAM8 by removing two levels from PAM8.

2.4.2 Simulation Setup

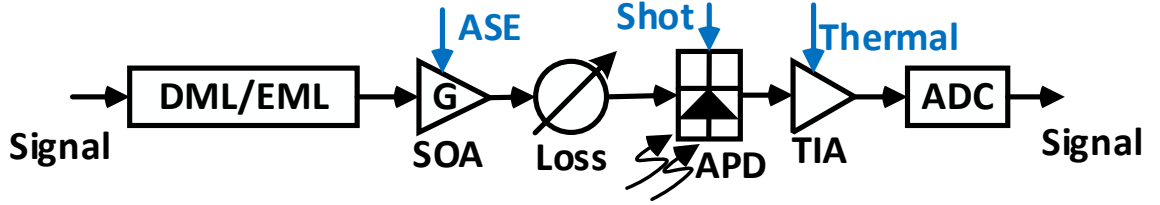


Figure 2. 16: Simulation setup of the flexible rate PON. ©IEEE 2020

Figure 2. 16 presents the 50-Gbaud simulation block diagram. The extinction ratio of an externally modulated laser or a directly modulated laser is set to 6 dB. The peak optical modulator output power is 4 mW and the minimum signal level is 1 mW. An SOA with 7 dB noise figure boosts the modulator output with variable gain (G). The SOA input SNR is 45 dB. The SOA output power (P_{avg}) is 6 dBm and SOA peak power threshold X_{th} is 10 dBm. The transmitter resolution is 8 bits and ADC resolution is 6 bits. Optical losses mainly arise from splitter loss and fiber loss. At the receiver side, the signal is detected based on square-law detection with signal-dependent shot noise and thermal noise. APD gain (M) is 8, responsivity is 0.7A/W @ $M=1$ and ionization factor is 0.13. TIA input-referred noise is 11 pA/sqrt (Hz). The signal bandwidth is assumed to be 25 GHz. The simulation is conducted under the assumption of linear operation range, perfect clock recovery and channel information, and no fiber dispersion.

2.4.3 Simulation Results with Flexible FEC Code Rates

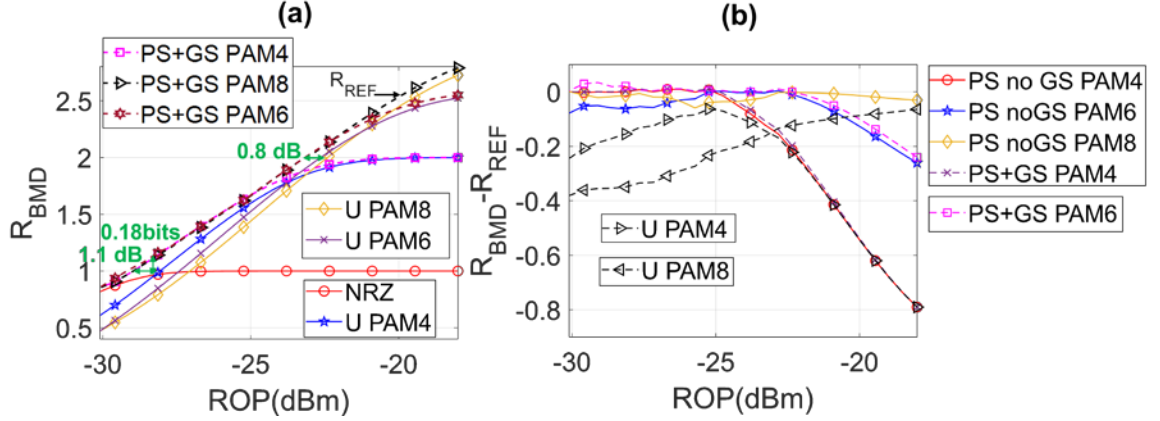


Figure 2. 17: Simulation results of signals without pairwise constraint. ©IEEE 2020

Firstly, we optimize the GMI without pairwise and NGMI threshold constraint. Figure 2. 17(a) presents the GMI of PS+GS signals. Compared with uniform PAM4 at GMI of 1 bits/symbol, PS signals exhibit 1.1-dB sensitivity gain and 0.18-bits GMI improvement. Compared with uniform PAM8 at GMI of 2 bits, PS PAM8 shows 0.8-dB sensitivity gain and 0.12- bits GMI improvement. We also plot the GMI difference between other settings and reference GMI (R_{REF} , GMI of PS+GS PAM8) in Figure 2. 17(b), where GS offers <0.05-bits GMI improvement compared to no GS settings.

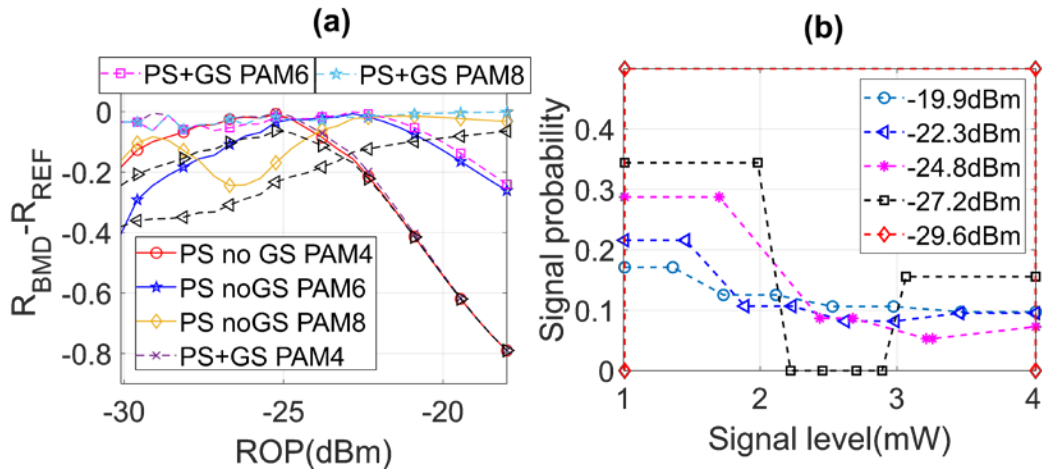


Figure 2. 18: Simulation results of signals with pairwise constraint. ©IEEE 2020

Next, we compare the signals with pairwise constraint by plotting the difference between GMI and R_{REF} as shown in Figure 2. 18(a). In this case NGMI has constraints defined by equation 23 (0.5 for PAM4, 2/3 for PAM6 and PAM8). Different from no pairwise setting, GS offers significant GMI improvement (up to 0.2 bits for PAM4, 0.4 bits for PAM6, 0.25 bits for PAM8) compared to no GS setting. No GS setting even performs worse than uniform signals with some lower ROPs. Meanwhile, with GS, the GMI nearly overlaps with R_{REF} , which means GS can compensate for the penalty from pairwise distribution. Figure 2. 18(b) presents the optimized PS and GS PAM8 results with pairwise constraint. PS tends to assign higher probabilities to symbols with less power, this reduces the modulator output power and requires an increase in SOA gain to maintain the same SOA output power. As the ROP decreases, the signal transitions to fewer levels by setting probabilities of some levels to 0 or merging some symbols into the same signal level. Moreover, GS tends to move symbols with higher probability away from one another. Table 2. 2 summarizes the SOA gain of pairwise GS+PS PAM8 versus ROP, which shows higher gain over the uniform signal (fixed at 2 dB).

Table 2. 2: SOA gain of PS+GS PAM8 with pairwise versus ROP ©IEEE 2020

ROP (dBm)	SOA gain (dB)
-18	2.45
-22	2.66
-26	2.65
-30	2

2.4.4 Simulation Results with a Fixed FEC Code Rate

In the case of pairwise constraint, we also sweep the NGMI threshold R_{th} considering a fixed FEC code rate implementation as shown in Figure 2. 19(a) and (b). PAM4 starts from $R_{th}=0.5$ while PAM8 starts from $2/3$. Without GS, as shown in Figure 2. 19(b), the signal starts to degrade from $R_{th}=0.68$. On the other hand, as shown in Figure 2. 19(a), the signal with GS shows degradation when R_{th} is larger than 0.78, and thus is less sensitive to the increasing of R_{th} compared to signal without GS.

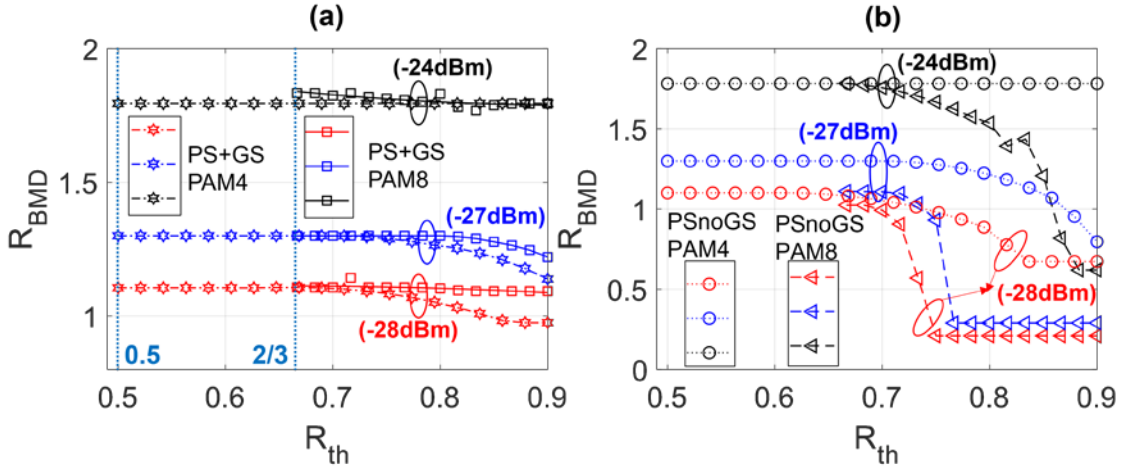


Figure 2. 19: GMI versus NGMI threshold R_{th} with pairwise constraint. © IEEE 2020

Next, we consider Monte Carlo testing with the IEEE 802.3ca LDPC code [102] with a fixed code rate (R_{fec}) of 0.84. According to the results of uniform PAM, the gap between NGMI and R_{fec} is $\Delta \approx 0.05$ at a post-FEC BER of 1×10^{-4} , which was chosen relatively high to make the simulation time feasible. Based on this observation, we perform the optimization with $R_{th}=R_{fec}+0.05=0.89$. With the optimized GS and PS settings, we run Monte Carlo testing to check the SE versus required ROP at the post FEC BER of 1×10^{-4} . Here the parity bits are inserted in the LSB of the DM output. LDPC decoding is realized by descrambling the LLRs to all zero-codewords at the receiver. The technique

relies on the assumption that the performance of LDPC decoders typically is independent of the exact values of the transmitted codewords and is consistent with the commonly applied technique of evaluating LDPC codes and decoders solely using all-zero codewords. Figure 2. 20 demonstrates that the proposed scheme always outperforms uniform signaling when the signal entropy is non-integer with continuous SE tuning. GS shows 3.8-dB and 2.3-dB sensitivity improvement compared to PS PAM8 without GS when SE=1 and 2 bits. PS PAM8 covers most of the ROP ranges and outperforms uniform PAM6 by 0.5 dB at SE=2.105 bits. PS PAM4 outperforms PS PAM8 by up to 0.1-bits/channel improvement from -25dBm to -24dBm.

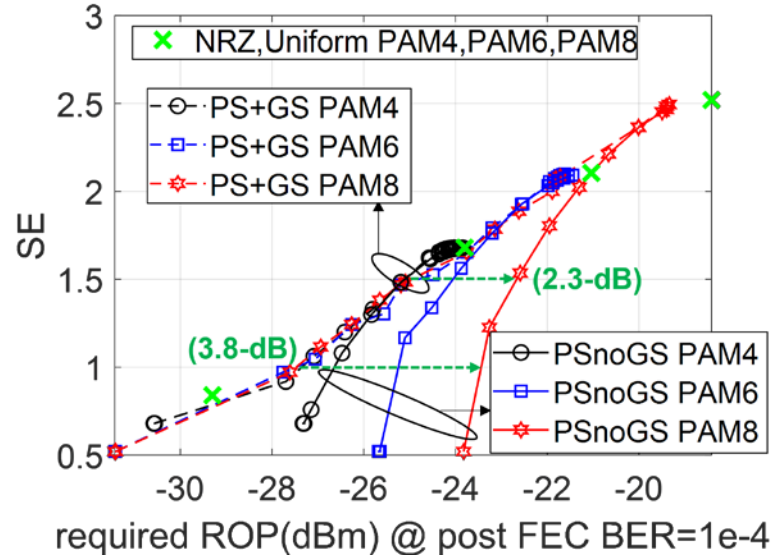


Figure 2. 20: Spectral efficiency versus required ROP with practical LDPC. ©IEEE 2020

2.4.5 Summary

In a summary, we have provided the first investigation of applying PS and GS to enable next generation flexible rate PON. PS and GS was studied and optimized

specifically for PON under the impact of different noises. With GS, the pairwise distribution can preserve the distribution from DM with comparable performance to the optimal results without constraints. In the case of flexible FEC code rate adjusting, the proposed scheme demonstrated up to 1.1-dB sensitivity gain and 18% net data rate improvement compared with uniform PAM4. PS PAM8 gives the optimal GMI among typical PON operating ranges. Monte Carlo simulation using practical LDPC also proved its feasibility to achieve flexible SE tuning with a fixed FEC code rate. The simulation results have been further verified through experimental demonstration [43].

CHAPTER 3. EFFICIENCY ANALYSIS AND DESIGN FOR HIGH-CAPACITY LINKS

In this chapter, we will address the efficiency issues in high-capacity communication systems. First, we will present the trade-off analysis between laser phase and signal performance for inter-data center and long-haul communications [60-62]. Then, we will investigate the cost-effective architecture design for intra-data center communications [65]. Finally, we propose a data-efficient nonlinearity correction algorithms for broadband communication systems [78, 79].

3.1 Impact of Frequency Jitter and Linewidth on Coherent Systems

This section shows the combined impact of laser jitter tones and laser linewidth on the DP coherent system through theoretical analysis, simulations, and experiments [60, 62].

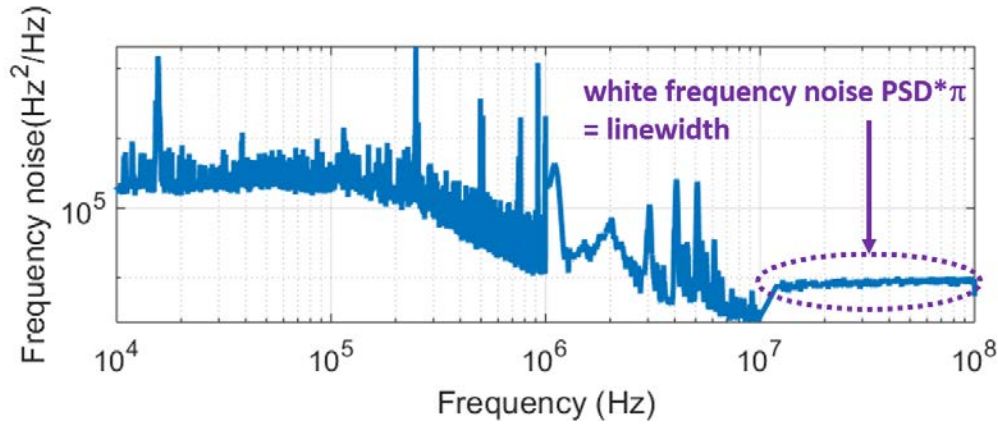


Figure 3. 1: Measured FN PSD in a coherent pluggable module. ©IEEE 2019

Figure 3. 1 shows the experimentally measured FN PSD in a state-of-the-art coherent pluggable module. The white noise (flat) region above 10MHz determines the laser linewidth by multiplying with π [51]. In the frequency region from 100 kHz to 10

MHz , there are multiple interfering tones due to the switching power supplies and circuit noise. The frequency jitter tones have a peak-to-peak frequency deviation Δf_{pp} ranging between $0.1 MHz$ and several MHz . Note that Δf_{pp} can be calculated from the power spectrum by carrying out $2\sqrt{2} \text{sqr}t(\cdot)$ operation on the power level of the sinusoidal interfering tone, which is the product of a tone amplitude in the FN PSD and its corresponding measurement resolution bandwidth. The LO phase noise incorporating not only the white frequency noise but also multiple sinusoidal jitter tones can be written as:

$$e^{j\phi_{LO}} = \exp \left(j \left(\sum_t \Delta f_{pp_t} / (2f_t) \cdot \cos(2\pi f_t t + \phi_t) + n_w(t) \right) \right), \quad (24)$$

where Δf_{pp} and f_t are the peak-to-peak frequency deviation and the sinusoidal jitter tone frequency, while ϕ_t corresponds to its initial phase, and $n_w(t)$ corresponds to the white frequency noise, which is modeled as a Wiener process.

3.1.1 Theoretical Analysis

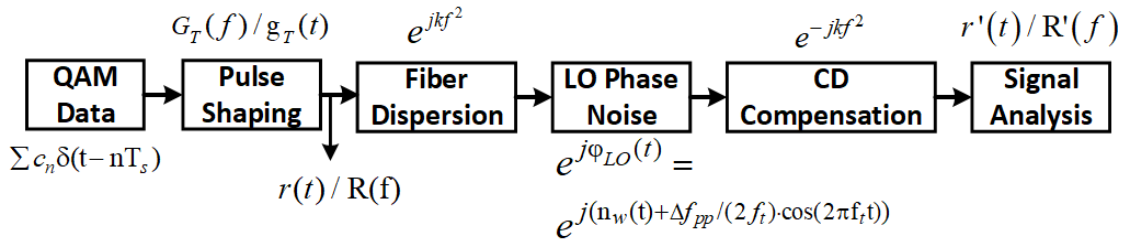


Figure 3. 2: Block diagram of the theoretical analysis with jitter tone. ©IEEE 2019

We perform a theoretical analysis using the block diagram shown in Figure 3. 2. Let the symbol period be T_s , and c_n be the any order of QAM symbols in the complex data plane, the incoming QAM data train is modeled as $\sum c_n \delta(t - nT_s)$. The signal then passes through an ideal Nyquist pulse shaping filter, whose interpolation function can be

expressed as $g_T(t) = \text{sinc}(t/T_s)$ to generate a band-limited signal. Considering that EEPN mainly comes from an LO rather than a transmitter in the post-CD compensation scheme[47], we assume a perfect transmitter laser and ignore the laser relative intensity noise (RIN) [47]. After fiber dispersion, the time domain of the demodulated signal influenced by EEPN can be modeled as [48]:

$$r'(t) = \int_{-\infty}^{\infty} X_{LO}(f_1) \cdot e^{-jkf_1^2} \cdot r(t - kf_1 / \pi) e^{j2\pi f_1 t} df_1, \quad (25)$$

where $X_{LO}(f_1)$ denotes the frequency response of the LO phase noise, which can be characterized as the Fourier transform of the phase fluctuation $e^{j\phi_{LO}}$. In equation 25, $k = \pi \cdot D \cdot L \cdot c / f_c^2$ is the accumulated dispersion factor, where D is the dispersion coefficient, c is light speed, f_c is the optical frequency. The transmitted signal after the Nyquist filter is $r(t) = \sum c_n \delta(t - nT_s) \otimes \text{sinc}(t/T_s) = \sum c_n \text{sinc}((t - nT_s)/T_s)$.

The theoretical analysis and numerical simulation have indicated that when only considering white frequency noise (laser linewidth), the square of EVM is proportional to the product of laser linewidth, fiber length, and baud rate [48]. In our work, a more general model is established by incorporating not only the white frequency noise but also multiple sinusoidal jitter tones, i.e., the LO phase noise which be written as:

$$e^{j\phi_{LO}} = \exp \left(j \left(\sum_t \Delta f_{pp_t} / (2f_t) \cdot \cos(2\pi f_t t + \varphi_t) + n_w(t) \right) \right), \quad (26)$$

where Δf_{pp} and f_t are the peak-to-peak frequency deviation and the frequency of a sinusoidal jitter tone while φ_t corresponds to its initial phase, and $n_w(t)$ corresponds to the white frequency noise which is related to laser linewidth.

We begin with deriving the general formula of a laser lineshape $S(f)$ when the laser phase noise is perturbed by a single sinusoidal jitter tone as a simplified case in equation 26. When considering only white frequency noise, $S(f)$ is modeled as a single Lorentzian distribution [48]. In the following analysis, we will show that when considering both white frequency noise and frequency jitter tone, $S(f)$ can be well approximated by a weighted combination of frequency-shifted Lorentzian distributions.

By using Bessel series expansion, we can obtain the spectrum of an LO phase noise with an interfering tone and white frequency noise as:

$$\begin{aligned}
S(f) &= \mathbb{E} |X_{LO}(f)|^2 \\
&= \iint_{\square^2} \mathbb{E} \left(e^{j(n_w(t) - n_w(u))} \right) e^{2\pi j(t-u)f} \\
&\quad \times \exp \left(j \frac{\Delta f_{pp}}{2f_t} (\cos(2\pi f_t t) - \cos(2\pi f_t u)) \right) dt du \\
&= \iint_{\square^2} \mathbb{E} \left(e^{j(n_w(t) - n_w(u))} \right) e^{2\pi j(t-u)f} \left(\sum_{n=-\infty}^{\infty} J_n \left(\frac{\Delta f_{pp}}{2f_t} \right) e^{-2\pi j n f_t} \right) \\
&\quad \times \left(\sum_{n=-\infty}^{\infty} J_n \left(\frac{\Delta f_{pp}}{2f_t} \right) e^{2\pi j n f_t} \right) dt du.
\end{aligned} \tag{27}$$

where $J_n(z) = \sum_{m \geq 0} \frac{(-1)^m}{(m+n)!} (z/2)^{2m+n}$ is the n -th order Bessel function. The expectation involving Wiener process can reach a closed form formula by using the characteristic functions of Gaussian random variables $\mathbb{E} \left(e^{j(n_w(t) - n_w(u))} \right) = e^{-\pi \cdot \Delta V \cdot |t-u|}$.

For each term of the summation in equation 27, by changing of variable $r = t + u, z = t - u$, we have:

$$\begin{aligned}
& \iint_{[0,T]^2} \mathbb{E} \left(e^{j(n_w(t) - n_w(u))} \right) e^{2\pi j(t-u)f} e^{-2\pi j m_1 f_t} e^{-2\pi j m_2 f_t} dt du \\
&= \frac{1}{2} \int_{-T}^T e^{-\pi \Delta \nu \cdot |z|} e^{2\pi j z f} e^{-\pi j z (n_1 + n_2) f_t} \int_{T-|z|}^{T+|z|} e^{-\pi j r (n_1 - n_2) f_t} dr dz \\
&\approx \begin{cases} \pi \int_{-T}^T e^{-\pi \Delta \nu \cdot |z|} e^{2\pi j z (f - n f_t)} dz, & n_1 = n_2 = n \\ 0, & n_1 \neq n_2 \end{cases}.
\end{aligned} \tag{28}$$

Therefore, we obtain:

$$\begin{aligned}
S(f) &\approx \pi \sum_{n=-\infty}^{\infty} J_n \left(\frac{\Delta f_{pp}}{2f_t} \right)^2 \int_{-\infty}^{\infty} e^{-\pi \Delta \nu \cdot |z|} e^{2\pi j z (f - n f_t)} dz \\
&= \frac{1}{2} \sum_{n=-\infty}^{\infty} J_n \left(\frac{\Delta f_{pp}}{2f_t} \right)^2 \frac{\Delta \nu}{(f - n f_t)^2 + \Delta \nu^2} \\
&= \frac{1}{2} \sum_{n=-\infty}^{\infty} w_n S_l(f - n f_t),
\end{aligned} \tag{29}$$

where $S_l(\cdot)$ is a Lorentzian-shaped function. Equation 29 shows that the LO phase noise spectrum which includes an interfering tone and white frequency noise is given by a weighted average of frequency-shifted Lorentzian distributions, with the weight at a frequency $n f_t$ given by $J_n \left(\frac{\Delta f_{pp}}{2f_t} \right)^2$.

The EVM caused by the phase to amplitude conversion in a receiver phase noise compensation block can be defined as the mean square error of the received signal after CD compensation normalized by the transmitted signal power [48]:

$$EVM^2 = \frac{\mathbb{E}(|r'(t)| - |r(t)|)^2}{\mathbb{E}(|r(t)|^2)}. \tag{30}$$

We calculate the mean square error of $|r'(t)|$ where $t = nT_s$. Since we have

$\mathbb{E}|r'(t)|^2 = \mathbb{E}|r(t)|^2$, using the Taylor expansion $\sqrt{1-x} = 1 - x/2 - x^2/8 + O(|x|^3)$ for small x ,

we get:

$$\begin{aligned} EVM^2 &= \frac{\mathbb{E}(|r(t)| - |r'(t)|)^2}{\mathbb{E}|r(t)|^2} = 2 - 2 \frac{\mathbb{E}(\sqrt{|r'(t)|^2 |r(t)|^2})}{\mathbb{E}|r(t)|^2} \\ &\approx \frac{1}{4\mathbb{E}|r(t)|^2} \mathbb{E} \left(\frac{|r(t)|^2 - |r'(t)|^2}{|r(t)|} \right)^2. \end{aligned} \quad (31)$$

By the definition of $r'(t)$ given in equation 25, we obtain:

$$\begin{aligned} EVM^2 &= \frac{1}{4\mathbb{E}|r(t)|^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathbb{E} \left(X_{LO}(f_1) \overline{X_{LO}(f_2)} X_{LO}(f_3) \overline{X_{LO}(f_4)} \right) e^{2\pi j(f_1 - f_2 + f_3 - f_4)t} \\ &\quad \cdot e^{-jk(f_1^2 - f_2^2 + f_3^2 - f_4^2)} \cdot \mathbb{E} \left(\frac{\left(r(t - \frac{kf_1}{\pi}) \overline{r(t - \frac{kf_2}{\pi})} - |r(t)|^2 \right) \left(r(t - \frac{kf_3}{\pi}) \overline{r(t - \frac{kf_4}{\pi})} - |r(t)|^2 \right)}{|r(t)|^2} \right) df_1 df_2 df_3 df_4. \end{aligned} \quad (32)$$

Note that phase noise terms at different frequencies have zero correlation. To make $\mathbb{E} \left(X_{LO}(f_1) \overline{X_{LO}(f_2)} X_{LO}(f_3) \overline{X_{LO}(f_4)} \right)$ non-zero, (f_1, f_2, f_3, f_4) must pair up into two pairs with appropriate conjugation. There are two possibilities: $(f_1 = f_2, f_3 = f_4)$, and $(f_1 = f_4, f_2 = f_3)$, leading to the following decomposition:

$$\begin{aligned} EVM^2 &= \frac{1}{4\mathbb{E}|r(t)|^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} S(f_1) S(f_3) \mathbb{E} \left(\frac{1}{|r(t)|^2} |r(t - \frac{kf_1}{\pi})|^2 |r(t - \frac{kf_3}{\pi})|^2 - |r(t)|^2 \right) df_1 df_3 \\ &\quad + \frac{1}{4\mathbb{E}|r(t)|^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} S(f_1) S(f_3) \mathbb{E} \left(\frac{1}{|r(t)|^2} \left| r(t - \frac{kf_1}{\pi}) \overline{r(t - \frac{kf_3}{\pi})} - |r(t)|^2 \right|^2 \right) df_1 df_3 \\ &= \frac{1}{4} (I_1 + I_2), \end{aligned} \quad (33)$$

For the first term I_1 , by doing mathematical calculation as shown in Appendix A.1, we obtain:

$$I_1 \approx a' \frac{k \cdot \Delta \nu}{T_s}, \quad (34)$$

Then we calculate $I_2 - I_1$ to obtain the second term I_2 as presented in Appendix A.2, direct calculation yields:

$$I_2 - I_1 \approx a' \frac{k \cdot \Delta \nu}{T_s} + 2 \int_{-\frac{T_s}{2k}}^{\frac{T_s}{2k}} \frac{k^2 f^2}{T_s^2} S(f) df. \quad (35)$$

Putting equation 34 and 35 together into equation 33:

$$\begin{aligned} EVM^2 &= \frac{1}{4}(I_1 + I_2) = \frac{1}{2}I_1 + \frac{1}{4}(I_2 - I_1) \\ &\approx \frac{a + 2a'}{4} \frac{k \cdot \Delta \nu}{T_s} + \frac{1}{2} \int_{-\frac{T_s}{2k}}^{\frac{T_s}{2k}} \frac{k^2 f^2}{T_s^2} S(f) df. \end{aligned} \quad (36)$$

For the second term $\int_{-\frac{T_s}{2k}}^{\frac{T_s}{2k}} \frac{k^2 f^2}{T_s^2} S(f) df$, it depends only on Δf_{pp} but not on f_t . Note that the

following identity holds true for series associated to Bessel J functions:

$$\sum_{n=-\infty}^{\infty} n^2 J_n(\beta)^2 = \frac{\beta^2}{2}. \quad (37)$$

Since $\Delta f_{pp}, f_t \ll T_s / k$, we have:

$$\int_{-\frac{T_s}{2k}}^{\frac{T_s}{2k}} \frac{k^2 f^2}{T_s^2} S(f) df \approx \int_{-T_s/2k}^{T_s/2k} \frac{k^2 f^2}{T_s^2} \left(\sum_{n=-\infty}^{\infty} w_n \cdot S_t(f - nf_t) \right) df$$

$$\begin{aligned}
&= w_0 \int_{-\frac{T_s}{2k}}^{\frac{T_s}{2k}} \frac{k^2 f^2}{T_s^2} \frac{\Delta \nu}{\pi(\Delta \nu^2 + f^2)} df + \sum_{n \neq 0} w_n \int_{-\frac{T_s}{2k}}^{\frac{T_s}{2k}} \frac{k^2 f^2}{T_s^2} \frac{\Delta \nu}{\pi(\Delta \nu^2 + (f - n f_i)^2)} df \\
&\approx w_0 \int_{-\frac{T_s}{2k}}^{\frac{T_s}{2k}} \frac{k^2 f^2}{T_s^2} \frac{\Delta \nu}{\pi(\Delta \nu^2 + f^2)} df + \sum_{n \neq 0} w_n \int_{-\frac{T_s}{2k}}^{\frac{T_s}{2k}} \frac{k^2}{T_s^2} \frac{(f^2 + n f_i f + n^2 f_i^2) \Delta \nu}{\pi(\Delta \nu^2 + f^2)} df \\
&\approx \int_{-\frac{T_s}{2k}}^{\frac{T_s}{2k}} \frac{k^2 f^2}{T_s^2} \frac{\Delta \nu}{\pi(\Delta \nu^2 + f^2)} df + \frac{k^2}{T_s^2} \sum_{n=-\infty}^{\infty} w_n n^2 f_i^2 \\
&= a \frac{k \cdot \Delta \nu}{T_s} + c \frac{k^2 \Delta f_{pp}^2}{T_s^2}.
\end{aligned} \tag{38}$$

Therefore, the EVM can be decomposed into two parts, representing the contribution from white noise and sinusoidal frequency jitter, respectively:

$$\begin{aligned}
EVM^2 &= \mathbb{E}(|r'(t)| - |r(t)|)^2 / \mathbb{E}(|r(t)|^2) = EVM_1^2 + \alpha \cdot EVM_2^2, \\
EVM_1^2 &\propto k \cdot \Delta \nu / T_s \propto L \cdot \Delta \nu / T_s, \\
EVM_2^2 &\propto k^2 \cdot \Delta f_{pp}^2 / T_s^2 \propto L^2 \cdot \Delta f_{pp}^2 / T_s^2.
\end{aligned} \tag{39}$$

In essence, this means that the integral under Lorentzian distribution (in equation 29) at carrier frequency leads to the term EVM_1^2 , representing the contribution from white noise; while each frequency-shifted, weighted term in equation 29 contributes to the EVM_2^2 term, representing the effect of sinusoidal frequency jitter. The relative weight α are independent of the fiber length, baud rate, and laser linewidth, but may vary with different DSP algorithms. The exact value of α can be obtained through numerical analysis.

The theoretical analysis above illustrates the dependence of EVM on baud rate, fiber distance, laser linewidth, and the peak-to-peak interfering tone amplitude. The result is suitable for different constellations. It is worth noting that, for higher order QAM, the required EVM would be more stringent, in this case, the system tolerance to laser linewidth and frequency jitter tone amplitude would be reduced.

3.1.2 Experimental and Simulation Setup

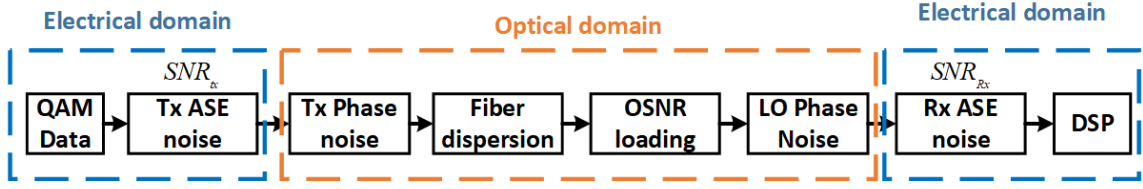


Figure 3. 3: Simulation setup to investigate jitter tone tolerance. ©IEEE 2019

The simulation procedure is shown in Figure 3. 3. Firstly, we measured the signal EVM at the transmitter (Tx) and receiver (Rx) in a back-to-back (BtB) setup to obtain EVM_{tx} and EVM_{total} , respectively, and convert them as SNR_{tx} and SNR_{total} , respectively, through $EVM^2 = 1/SNR$ [101]. The transmitter SNR_{tx} includes the EDFA's and DAC's amplified spontaneous emission (ASE) noise. The receiver SNR (SNR_{rx}), which accounts for the intradyne coherent receiver (ICR) and analog to digital convertor (ADC) induced SNR degradation, can be obtained from $\frac{1}{SNR_{total}} = \frac{1}{SNR_{tx}} + \frac{1}{SNR_{rx}}$. Secondly, the transmitter laser phase noise is based on a 35kHz linewidth, and the receiver LO phase noise is based on a certain linewidth with jitter tone. Thirdly, the fiber link is modeled as $g(f) = \exp(-j\frac{\pi \cdot c \cdot L \cdot CD \cdot f^2}{(f_c - f)^2})$ [103], where f_c is the carrier frequency, L is the fiber length, and CD is the fiber dispersion coefficient. CD is set as 17 ps/(nm×km) in simulation. Finally, ASE noise loading was applied to the received signal to vary the received optical signal-to-noise ratio (OSNR).

Figure 3. 4 shows the experimental setup of the DP-16-QAM coherent optical system. Four uncorrelated data sequences were applied with pre-equalizer with 11 taps and then were loaded to four 64 GSa/s DACs operating at 1 Sa/symbol. The signal is then

modulated by a coherent driver modulator (CDM), whose 3dB bandwidth is 40 GHz and V_{pi} is 2V. A CD emulator with an 8-dB insertion loss was used to emulate a total CD of 450 or 900 km standard SMF. Comparing to using SMF spools and in-line EDFAs, the CD emulator completely removes the concern of mixing fiber nonlinearity. At Rx, a PM and a low-speed AWG were used to modulate the LO with a sinusoidal tone by loading a sine wave to the AWG.

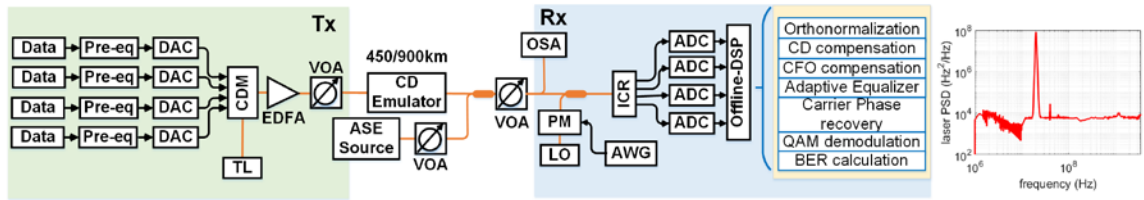


Figure 3. 4: Experimental setup to investigate jitter tone tolerance. ©IEEE 2019

As presented in the inset of Figure 3. 4, after PM, the laser frequency noise PSD exhibits a sharp peak at 20 MHz with a peak-to-peak amplitude (Δf_{pp}) equals 25.77 MHz. By changing the driving voltage of the PM and the period of the loading sine wave, we can sweep Δf_{pp} and frequency (f_i) of the sinusoidal tone in the experiment. After the ICR, a 4-channel 80-GSa/s real-time scope captures the data at 1.25 Sa/symbol. In the offline-DSP, the data is resampled to 2 Sample/symbol, orthonormalized using Gram-Schmidt orthogonalization procedure, followed by CD and CFO compensation, a 64-tap 4x2 butterfly MIMO AE, and blind-phase search CPR. The simulation and experiment share the same Rx DSP procedure.

3.1.3 Experimental and Simulation Results

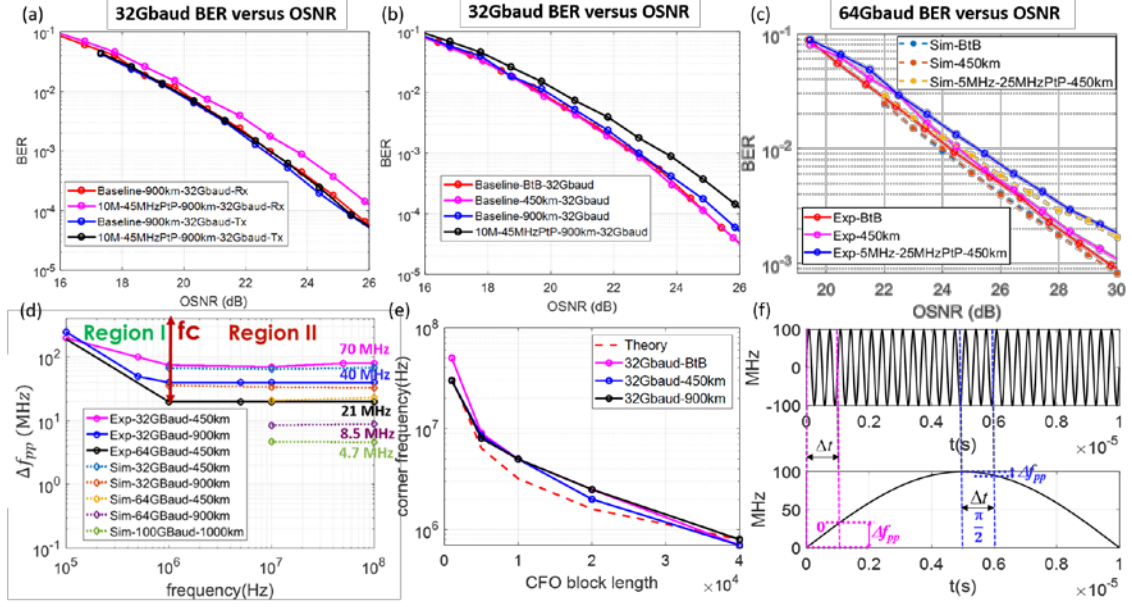


Figure 3. 5: Experimental and simulation results with the impact of jitter tone. (rec. opt. power= -8dBm). ©IEEE 2019

As shown in Figure 3. 5(a), firstly, we compare the BER versus OSNR performance between loading frequency jitter tone at the transmitter laser and LO. We can observe that a 10-MHz tone with $\Delta f_{pp} = 45$ MHz at LO side induces 0.8-dB OSNR penalty at the BER of 1×10^{-2} while the curve of the same tone at the transmitter side shows no penalty. This is because the transmitter laser phase noise has little impact on the system performance [48]. Therefore, we will focus on effect due to a frequency jitter tone imposed on an LO only. Figure 3. 5(b) and (c) illustrate the BER versus OSNR after adding a jitter tone to an LO with different baud rates. The results indicate that a higher baud rate system would be more sensitive to a jitter tone. The dashed lines (simulation results) match well with the experimental results.

Figure 3. 5(d) presents the measured LO frequency jitter mask, which is defined as the allowable Δf_{pp} versus the jitter frequency f_i with a certain penalty threshold. We set the

threshold as a 0.5-dB rOSNR penalty at a BER of 1×10^{-2} . Note that in high frequency (e.g., $> 1\text{MHz}$), the jitter mask is independent of the tone frequency. A tighter mask is required when the baud rate is higher, or the transmission distance is longer. For example, for a 64-Gbaud transmission, the maximum allowable Δf_{pp} is reduced from 21 MHz to 8.5 MHz when the distance varies from 450-km to 900-km. The mask floor is further reduced to 4.7 MHz for a 100 Gbaud DP-16QAM signal over 1000 km. Based on the measured Δf_{pp} range in Figure 3. 1, we can say that the impact of frequency jitter tone is of particular importance for 100 Gbaud/DP-16QAM and beyond, and for over 1000km. The frequency jitter mask is divided into two regions by a corner frequency f_c , which is inversely proportional to the block length (in time) of the CFO compensation. As shown in Figure 3. 5(e), the corner frequency is roughly proportional to $1/(\text{Symbol block length} \cdot T_s)$. The block length (in number of symbols $\times T_s$) corresponds to the CFO estimator observation time. In CFO estimation, we usually utilize a sufficient long block size to get a more accurate CFO value. However, the longer the CFO block length, the longer the observation time of phase noise [53], which in turn causing a reduced corner frequency and a tighter jitter mask. The slow frequency drift in Region I does not affect laser line shape too much and is handled by the CFO compensation section, while the faster frequency jitter in Region II affects the laser line shape and results in phase variance as well as EEPN [49]. Figure 3. 5(f) illustrates that the frequency deviation in Region I is not only determined by Δf_{pp} but also related to initial phase when the jitter tone does not complete a period in the observation time [49]. The faster frequency jitter such as the upper curve can complete at least one period for the whole observation time while the slower frequency jitter (lower curve) may not complete a complete cycle depending on the initial phase. For instance, the

π initial phase induces larger Δf_{pp} than $\pi/2$ initial phase in Figure 3. 5(f). In our experiment, the observation time is around 1 μ s. In Region I, we have swept the initial phase of tones and chosen the worst case Δf_{pp} to ensure the tightest Δf_{pp} tolerance. For BtB transmissions, the tolerance in Region II is dominated by CPR capability. After CD emulator, the tolerance mask is much tighter due to EEPN. Thus, the frequency jitter tolerance in Region II is obviously much tighter than that of Region I. In the following analysis, we focus on the system impact of frequency jitter tones in Region II because that is where the worst performance occurs.

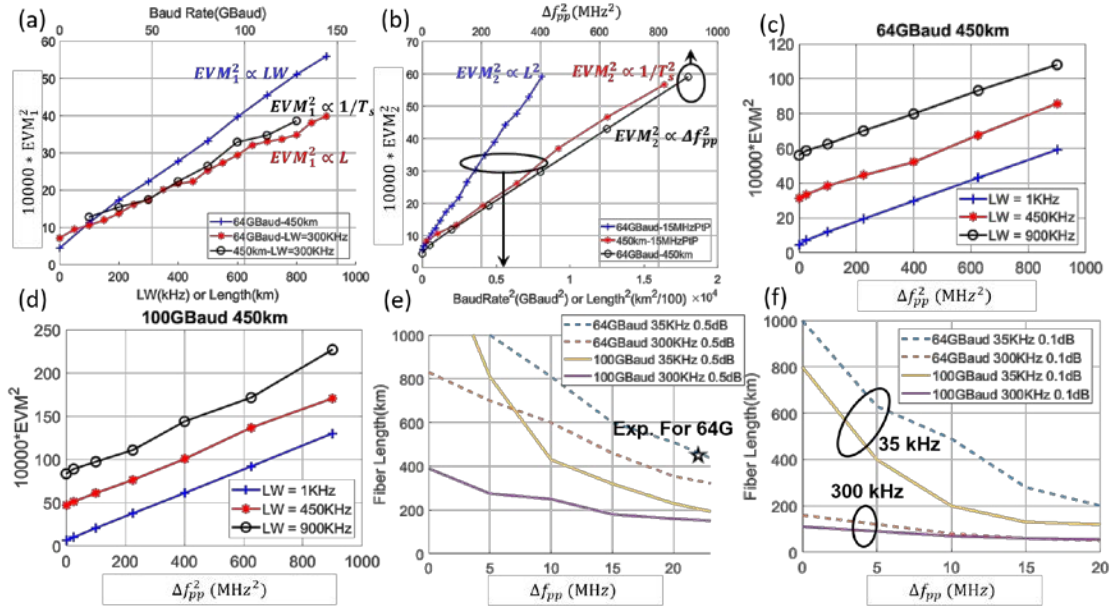


Figure 3. 6: Simulation results (for 64Gbaud/DP-16QAM signals) with single jitter sinusoidal tone to validate theoretical conclusion. ©IEEE 2019

Then we validate equation 39 as shown in Figure 3. 6. Figure 3. 6(a) demonstrates the proportionality of the EVM_1^2 caused by the white frequency noise by setting EVM_2^2 to zero. Figure 3. 6(b) demonstrates the proportionality of EVM_2^2 caused by the frequency jitter by setting EVM_1^2 to a small value with laser $\Delta\nu$ equals to 1kHz. EVM_2^2 versus L^2 is

obtained with the $\Delta f_{pp} = 15$ MHz, EVM_2^2 versus baud rate² is obtained with $\Delta f_{pp} = 15$ MHz and $L = 450$ km, and EVM_2^2 versus Δf_{pp}^2 is presented with 450-km transmission. Figure 3. 6(c) and (d) show the simulation results of EVM^2 versus Δf_{pp}^2 at a transmission distance of 450 km for 1, 450, and 900 kHz laser linewidths in 64-Gbaud/DP-16QAM and 100-Gbaud/DP-16QAM systems, respectively. These results prove that EVM_1^2 and EVM_2^2 are additive since the three curves are parallel and show even gaps between each other. For 100-Gbaud/DP-16QAM system, a change of 450 kHz in linewidth is the same as a change in Δf_{pp}^2 of approximately 300 MHz². Figure 3. 6(e) and (f) show the achievable fiber distance versus Δf_{pp} for 64 and 100 Gbaud DP-16QAM using two laser linewidths (35 kHz and 300 kHz), with an OSNR penalty of 0.5dB and 0.1dB at a BER of 1×10^{-2} . Experimental results, shown by the star symbol, are also included, which match well with the simulation results. For the typical range of Δf_{pp} of 0.5 MHz to several MHz, the 35 kHz linewidth laser exhibits much longer fiber reach than that of the 300 kHz linewidth laser. Moreover, the higher the baud rate is, the shorter the transmission distance is. It implies when the baud rate is beyond 100 Gbaud, laser linewidth close to or even lower than 35 kHz would be required.

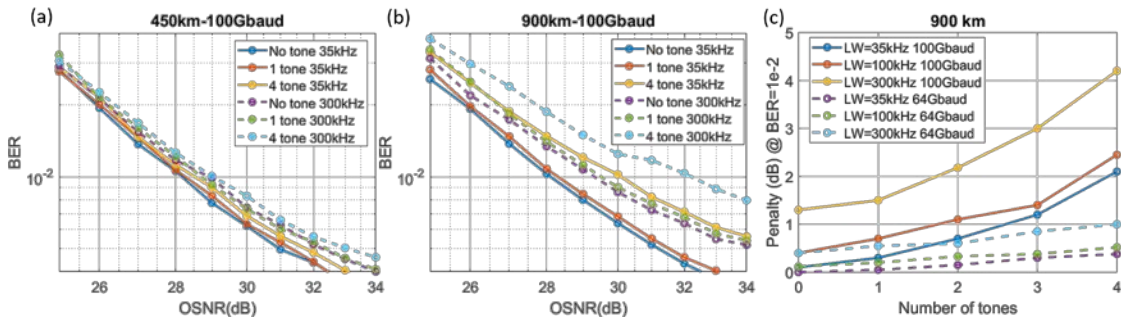


Figure 3. 7: Simulation results of the impact of multiple tones. ©IEEE 2019

In practice, multiple frequency jitter tones could occur (as shown in Figure 3. 1). Therefore, we performed simulations of multiple tones with random initial phases. All the tones are set as the same Δf_{pp} of 5 MHz. Figure 3. 7(a) and (b) illustrate the BER performance in 450-km and 900-km 100-Gbaud DP 16QAM transmission under the impact of multiple tones. For 450-km and a single tone, 300-kHz linewidth exhibits ~0.6-dB penalty at the BER of 1×10^{-2} compared with that of a 35 kHz linewidth, while the penalty from the additional four jitter tones are not significant (around 0.25 dB). When the distance is increased to 900 km, the penalty from the additional 4 tones becomes higher. With four tones (f_t at 1MHz, 3MHz, 6MHz, and 10MHz), the additional penalty from tones is 2 dB for 35kHz and 2.5 dB for 300kHz while for a single tone the penalty is small (0.15 dB). It indicates that the existence of multiple jitter tones would result in a tighter frequency jitter mask. Figure 3. 7(c) presents the result of the OSNR penalty (relative to back-to-back case without any interfering tone) at a BER of 1×10^{-2} with respect to the number of jitter tones. With the increasing of the number of jitter tones, the penalty also increases. It is worth noting that the increasing of penalty is not linear but a bit exponential for the case of 100Gbaud. Moreover, the combined effect of higher laser linewidth and jitter tones would introduce more penalty in high baud rate transmission.

3.1.4 Summary

In a summary, through theoretical, simulations, and experiments, we have found that, while EVM^2 due to EEPN is proportional to (baud rate \times LO linewidth \times transmission distance) when considering only the laser linewidth effect, it is proportional to (baud rate \times frequency jitter tone peak-to-peak frequency drift \times transmission distance)² when

considering the sinusoidal frequency jitter tone effect. The latter is an important factor to consider for a transmission system performance because the sinusoidal jitter tones often occurs in a pluggable transceiver or line-card due to unavoidable switching power supplies and other circuit noise. As a result, it is expected that a narrower laser linewidth would provide a higher margin for LO frequency jitter tolerance. It indicates the importance of setting up a jitter tolerance mask to qualify tunable lasers based on different baud rates and fiber lengths.

3.2 Impact of Laser Flicker Noise on Ultra-Baud Rate Coherent Systems

As indicated in Chapter 1, flicker noise in a DBR tunable laser, on the other hand, has been shown to cause significant coherent system penalty even without any transmission fiber [54, 55]. However, the examined flicker noise level was excessively high due to the current injection-induced carrier density fluctuation in the phase sections. This excessive frequency noise can be reduced by using thermal tuning instead [104]. In this section, we have used a thermally-tuned DBR LO whose flicker noise PSD closely follows that of an industry forum (OIF 400ZR) to investigate the impact of laser flicker noise and linewidth on ultra-baud rate coherent systems.

3.2.1 Experimental Setup

Figure 3. 8 shows the experimental setup. The transmitter is a 193.5 THz ECL, and three types of LOs were used: an ECL ($\Delta\nu=47\text{kHz}$), a DBR ($\Delta\nu=210\text{KHz}$), and the same DBR with additional injected flicker noise to make its FN PSD match with that of the industry forum (OIF 400ZR mask). The latter was achieved by using a PM and a low-speed AWG to modulate the phase of the DBR LO. The FN PSD for the DBR (PSD A) and DBR

with additional flicker noise (PSD B) are shown in Figure 3. 9. The PSD B has additional flicker noise between 1 and 70 MHz, which matches with the OIF mask well. For the area between 70 and 100MHz, PSD B falls slightly below the OIF Mask due to bandwidth limitation of the PM. At the Tx, the data were loaded to an InP CDM through a 120GS/s AWG whose 6dB bandwidth is 47 GHz. The 6-dB bandwidth of the CDM is 49 GHz. Pulse shaping with a roll-off factor of 0.2 and pre-equalization via 100 taps were employed. The maximum transmission distances for 64Gbaud/DP-64QAM and 96 Gbaud/DP-32QAM were 450km (a CD emulator having an insertion loss of 8 dB) and 400km (SSMF with a span length of 50km), respectively. The input power for each fiber span is set at 0 dBm/channel.

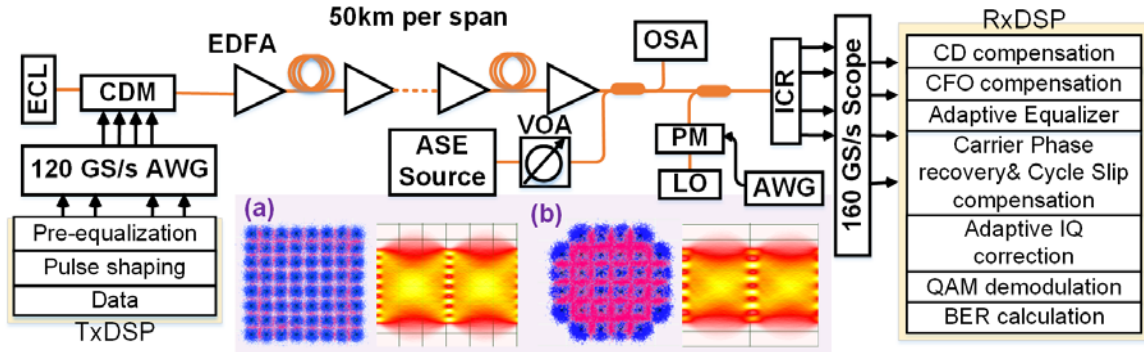


Figure 3. 8: Experimental setup for flicker noise investigation. ©OSA 2020

At the receiver side, an InP ICR with a 6-dB bandwidth of 51GHz, and a 160-GS/s scope were used. The input signal and LO power to the ICR were -6dBm and +12.5 dBm, respectively. In the Rx DSP, the signal was resampled to 2 samples per symbol, followed by a CD and a CFO estimation and compensation. A 4×2 butterfly AE is then employed to perform polarization de-multiplexing, residual CD, and residual clock frequency error. The AE uses a training sequence for pre-convergence and LMS for tracking. Since the AE is

implemented in the frequency domain, the taps and carrier phase are updated with a time interval equals to $n\text{Taps}/(2B_s)$. After the AE, the blind phase search (BPS) algorithm [9] with 64 test angles and 65 symbols sliding window averaging is used to refine the carrier phase estimation. Note that the BPS algorithm gives a better OSNR performance than a pilot-aided phase recovery algorithm [105, 106], e.g., 1 in 32 pilot injection in OIF 400ZR Implementation Agreement. The phase noise-induced OSNR penalty would be higher than what we report here if a pilot-aided algorithm is used. After the CPR, an adaptive IQ correction is used to compensate for residual IQ crosstalk [89].

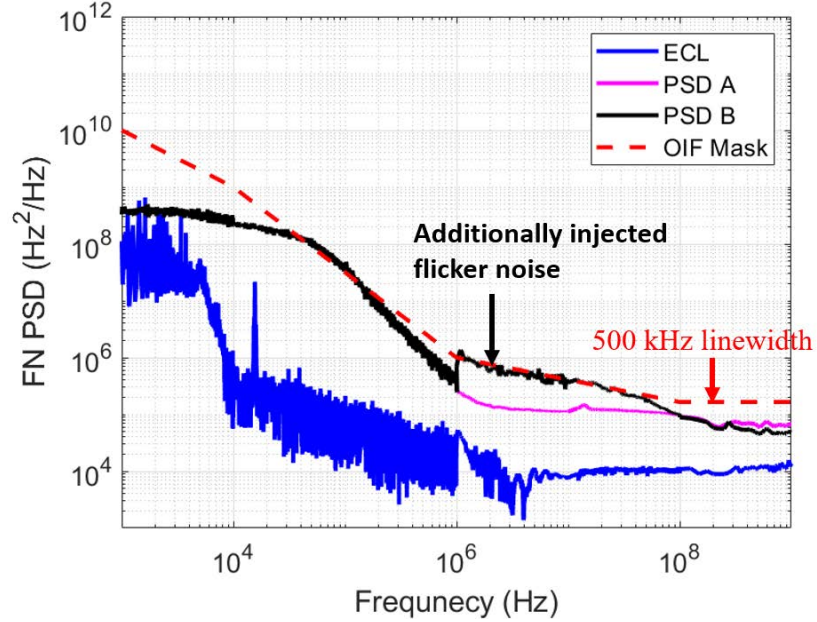


Figure 3. 9: FN PSD of different lasers in the experiment. ©OSA 2020

3.2.2 Experimental Results

First, we investigate the FN PSD impact on 64Gbaud or 96Gbaud DP 16-QAM signal. Figure 3. 10 presents the experimental results. The tap numbers for all types of LOs are 64. For BtB transmission, there is no OSNR penalty for different LOs. However, EEPN

becomes more stringent with the increasing of fiber distances. In Figure 3. 10(a), we can observe that, different types of LOs induce little OSNR degradation (~ 0.1 to 0.2 -dB) in 64Gbaud 450-km transmission. However, in case of 900-km transmission, compared with ECL, PSD A shows 0.3 -dB OSNR degradation at the BER of 1×10^{-2} while PSD B shows 0.5 -dB penalty. For higher baud rate (96 Gbaud) as shown in Figure 3. 10 (b), compared with ECL at the BER of 1×10^{-2} , PSD B induces 0.5 -dB OSNR penalty while PSD A shows 0.3 -dB OSNR penalty with a 400-km fiber dispersion.

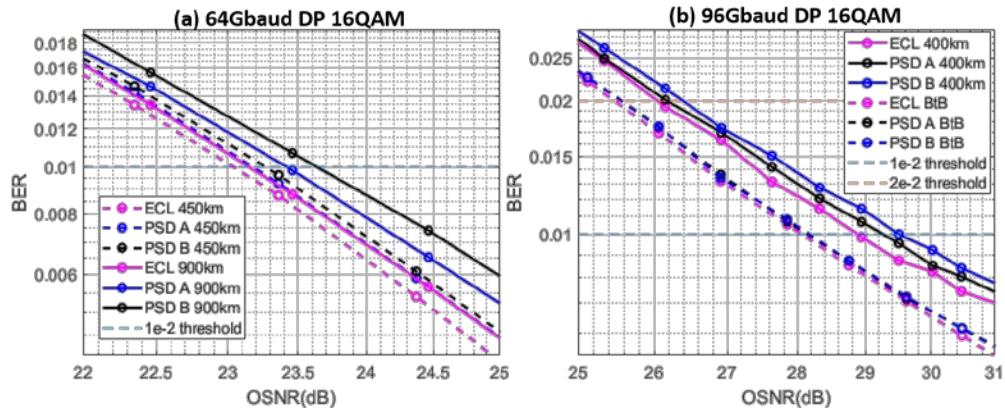


Figure 3. 10: BER versus OSNR of 16QAM signal. ©OSA 2020

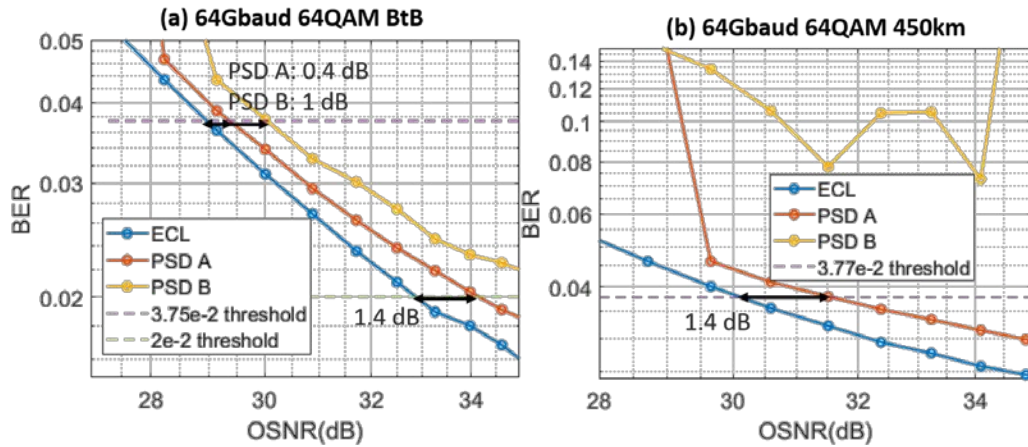


Figure 3. 11: BER versus OSNR of 64QAM signal (a) in BtB case and (b) with 450-km transmission. ©OSA 2020

Different from 16QAM signals with the same AE tap number for different types of LOs, the 64-Gbaud/DP-64QAM signal is very sensitive to LO phase noise. Therefore, different AE tap numbers were used for different types of LOs to optimize the system performances. A larger AE tap number such as 64 has more symbols in a processing block, so more AWGN can be removed due to additional averaging. However, a higher order of modulation is more sensitive to phase fluctuations and consequently requires a smaller tap number and processing block to better track the phase variation. For the LO with a frequency noise PSD A and PSD B, the BER performances become unstable when we employ larger tap number (e.g., 64), especially in case of lower OSNR. Therefore, the optimized tap numbers were 64, 32, and 16 for the ECL LO, the LO with FN PSD A and B, respectively. Figure 3. 11(a) shows that for 64Gbaud/DP-64QAM, significant OSNR performance difference exists among the three types of LO even in BtB condition. In comparison to an ECL, the LO with PSD A incurred 0.4-dB OSNR penalty at a BER of 3.77×10^{-2} [107], while the LO with PSD B incurred a penalty of 1 dB. An FEC with a smaller overhead would cause higher OSNR penalties. For example, PSD A and B incurred 1.4-dB and $\gg 2$ -dB OSNR penalty at a BER of 2×10^{-2} , respectively. These penalties mainly come from different tap numbers, hence different levels of AWGN removing. Next, the EEPN effect was investigated by using a 450-km CD emulator in Figure 3. 11(b). The LO with a frequency noise PSD B fails with 450-km fiber dispersion while the LO with a frequency noise PSD A incurred a 1.4-dB OSNR penalty at a BER of 3.77×10^{-2} .

Figure 3. 12 presents the estimated phase fluctuation over time in different LOs at an OSNR of 35 dB in the 64G/DP-64QAM experiment. The effective laser linewidth can be obtained via the phase difference variance through $LW_{eff} = var\langle\Delta\phi\rangle/(2\pi\tau)$, where

$var\langle\Delta\phi\rangle$ denotes the phase difference variance, and τ is the time interval between two phases. For a fair comparison, we used the same time interval ($\tau = 64/(2B_s)$) to calculate the effective linewidth. In both BtB and after 450km CD emulator, ECL exhibits smooth curves, while PSD A and B exhibit significant phase fluctuations. The ECL effective linewidths before and after the 450km SMF are the same, and match well with the actual measurement. For PSD A, after the 450km CD emulator, the effective linewidth increases from 192.7 kHz (which matches well with measurement) to 329.3 kHz. The introduction of extra frequency flicker noise in PSD B causes the effective linewidth to change significantly from 458.8 kHz (which is much higher than the measured 210kHz linewidth, due to the additional flicker noise) to 610 kHz after 450km CD emulator.

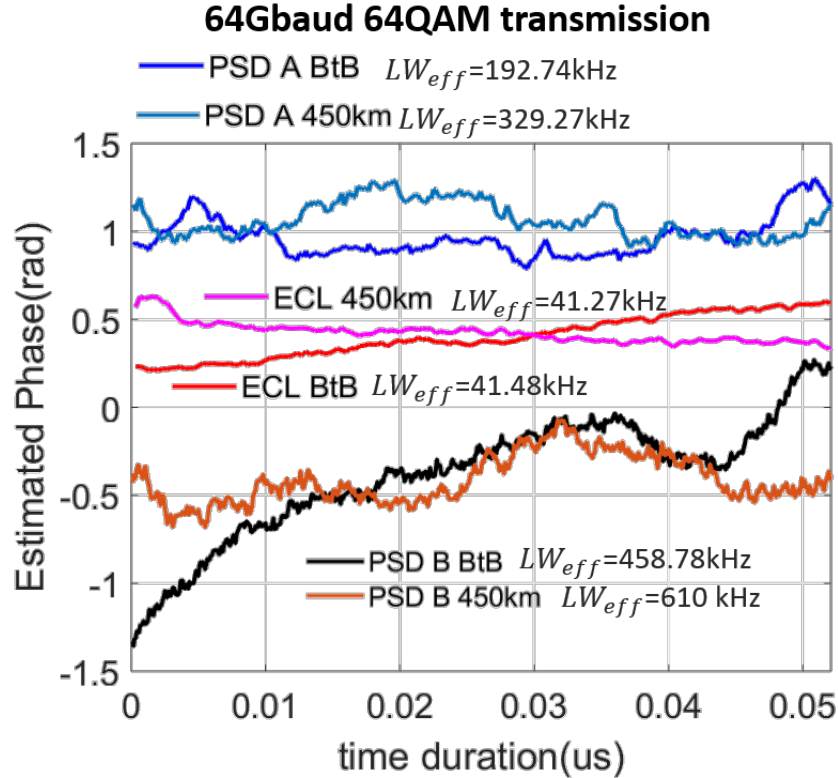


Figure 3. 12: Estimated phase fluctuations versus time duration with different effective linewidths. ©OSA 2020

Note that in a real-time system, the symbol rate and ADC sampling frequency are higher than the DSP clock frequency, thus the carrier phase recovery algorithm needs to resort to parallel processing or pilot-aided algorithm [106], which in turn makes the phase noise tolerance significantly reduced compared to our results presented here [108].

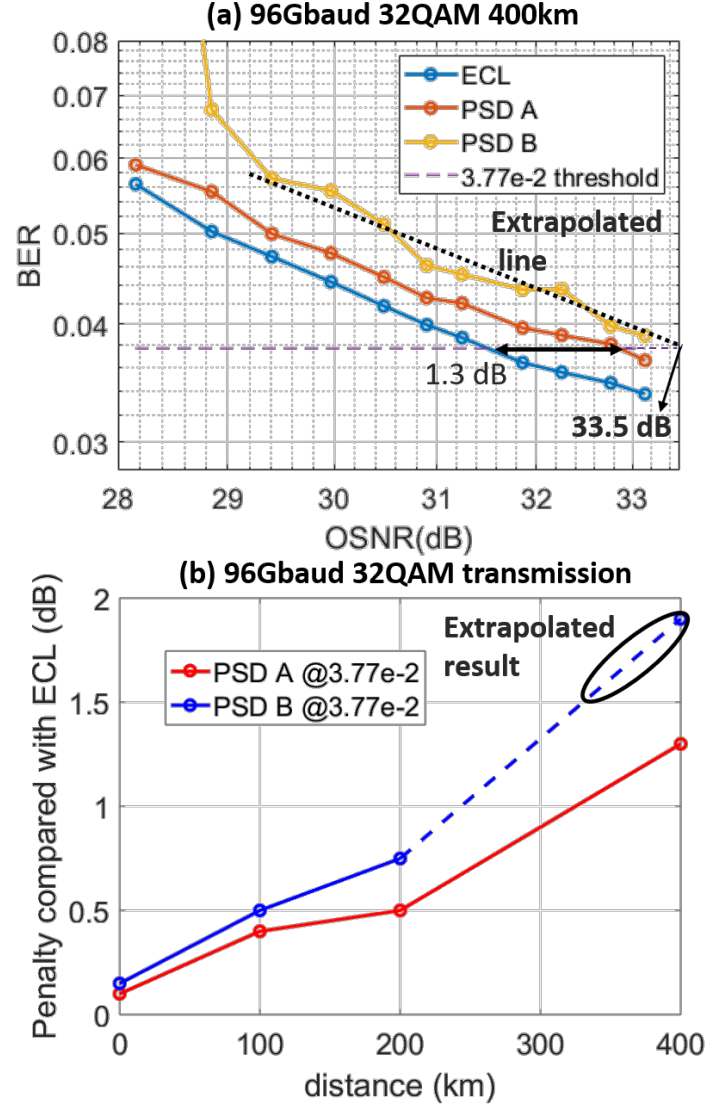


Figure 3. 13: Experimental results of 96-Gbaud 32QAM signal. ©OSA 2020

In the 96-Gbaud/DP-32QAM transmission experiment shown in Figure 3. 13(a) and (b), the signal is less sensitive to phase variation than 64-Gbaud/DP-64QAM signal.

In a BtB setup, the BER versus OSNR performance becomes very close for all three types of LO. However, the EEPN-induced OSNR penalty increases with the transmission distance. With 400-km SSMF transmission, 1.3 and 1.9-dB OSNR penalties are observed when using PSD A and PSD B, respectively. Figure 3. 13(b) summarizes the OSNR penalties caused by PSD A or B when compared to the ECL at different SSMF lengths. Both types of DBR exhibit ≤ 0.5 dB OSNR penalty when the reach is under 100km. However, this penalty increases with fiber length, and the penalty due to the DBR LO with a PSD B increases much faster than that of the DBR LO with a PSD A.

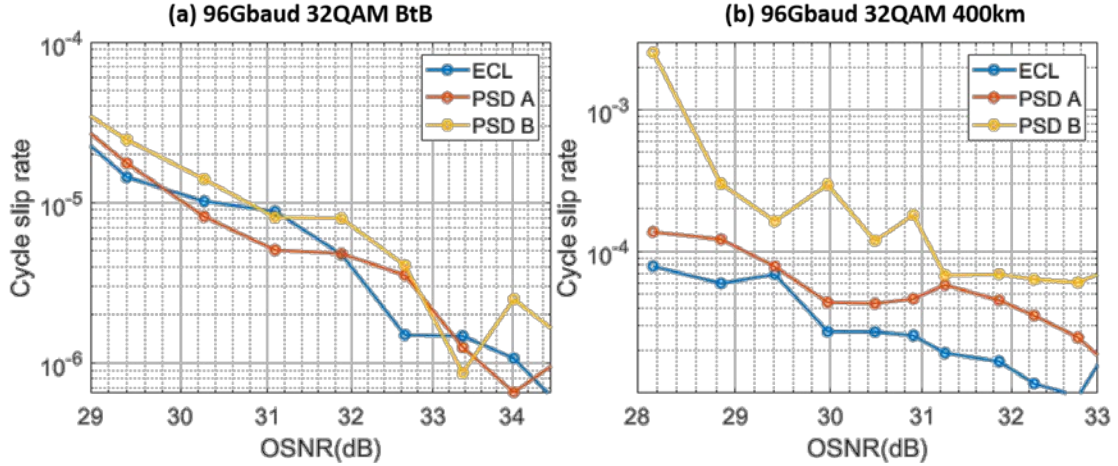


Figure 3. 14: Cycle slip rate versus OSNR of 96-Gbaud 32QAM signal (a) in BtB case and (b) with 400km fiber. ©OSA 2020

Moreover, we further investigate the cycle clip rate performance in 96 Gbaud 32-QAM systems. Different from the significant cycle slip observed in previous report for BtB case [54], with additionally flicker noise and higher linewidth, the cycle clip rates of PSD A and PSD B are still similar to that of ECL as shown in Figure 3. 14(a). With the fiber distance increases to 400 km, the cycle slip rate of PSD B increases more significantly than PSD A and ECL. The additional flicker noise or higher linewidth would induce more cycle

slips with the increasing of fiber distance. This is consistent with our observations in Figure 3. 12.

3.2.3 *Summary*

In summary, based on an OIF 400ZR laser frequency noise mask, we have investigated the impact of laser linewidth and flicker noise on coherent systems with different QAM orders, baud rates and transmission distances. We have found that, despite the 16-QAM system is less affected by phase fluctuations and EEPN, the 64-QAM system is sensitive to laser linewidth and flicker noise even in a BtB case. Furthermore, the flicker noise above 1MHz, despite its low amplitude, forbids the 450km system to work. The 96Gbaud DP 32-QAM system is sensitive to laser linewidth and flicker noise only beyond 200km, and the OSNR penalty of using a DBR laser in comparing to the case of an ECL laser was up to 1.9 dB at 400km. Note that in practice, a real-time DSP most likely would use a simplified carrier recovery algorithm (for lower power consumption) such as pilot symbol-assisted algorithm, which cannot achieve the superior performance of BPS carrier recovery, and therefore a higher OSNR penalty than what we have observed would be incurred. The EEPN penalty is also found to be independent of the roll-off factor of pulse shaping [48]. Consequently, for future metro optical systems which transport beyond 600G/λ, a tighter flicker noise and linewidth mask than what is currently defined in OIF 400ZR is needed for both transmitter and LO.

Considering that EEPN is hard to be mitigated through DSP, we have performed a trade-off analysis between laser phase noise and system performance for long-reach coherent optical communications. Based on different system settings, we can choose the

most cost-effective laser phase noise mask. In the next section, we will design a cost-effective coherent system for short-reach applications.

3.3 Cost-Efficient Coherent System Design

While conventional coherent systems are deployed in inter data center and long-haul applications as shown in Section 3.2 and 3.3, they are not cost-efficient for the future intra-data center applications. A self-homodyne (SHD) coherent optical transceiver has a great potential to reduce the cost and power consumption of a conventional coherent transceiver. This is due to that, by sending both the modulated signal and a copy of the transmitter laser (i.e., a remote LO) to a receiver through two separate short optical fibers with similar lengths, the portion of DSP related to carrier frequency offset and carrier phase recovery can be removed. While at the same time, if a silicon-photonics-based adaptive polarization controller [109] can be built into a coherent receiver for the signal and/or remote LO, the portion of DSP related to polarization de-multiplexing can also be removed [6, 14, 63, 110, 111]. Consequently, at a lower baud rate of 50~60+ Gbaud and 23 Gbaud, a “MIMO-free” coherent DSP was claimed possible [6, 111], except that the in-/quadrature- (I-Q) phase rotation due to path-length mismatch between the signal and remote LO still needs to be corrected.

In this dissertation, assuming a low-latency FEC based BER threshold, we experimentally demonstrate a single-carrier 800Gb/s SHD coherent link by using 120 Gbaud/DP-16QAM modulation. The link was run at C-band to support the future high-capacity dense WDM links required for artificial intelligence and machine learning.

3.3.1 Experimental Setup

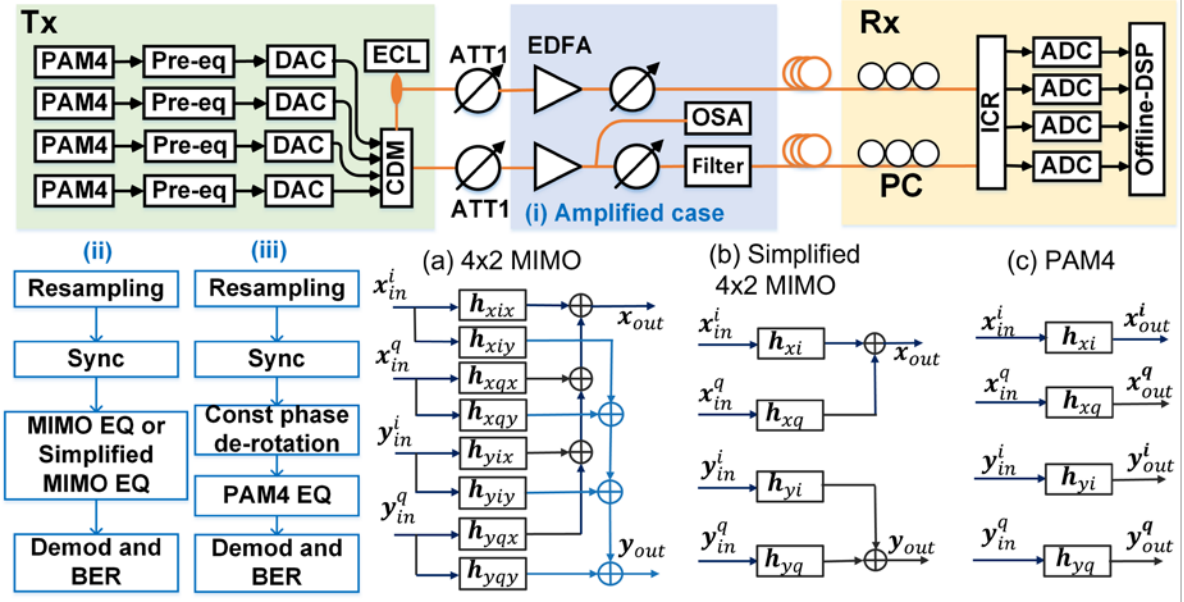


Figure 3. 15: Experimental setup and DSP block diagrams of the SHD coherent link. (i) is used in amplified case. ©OSA 2022

Figure 3. 15 shows the experimental setup and DSP flow. At the transmitter side, four uncorrelated PAM4 sequences were pre-equalized and loaded into a 120GS/s 70-GHz DAC operating at 1 Sa/Symbol and passing through a CDM, whose 3-dB bandwidth was 61 GHz. A 16.5 dBm external cavity laser is split into two paths using a 50:50 polarization maintaining coupler, which has a linewidth of 35 kHz and operates at a frequency around 193.1 THz. One path is used for Tx signal modulation and the other path for the remote LO through separate fibers to the coherent receiver. The optical power into CDM was 12.5 dBm and the remote LO launched into the transmission fiber was also 12.5 dBm. The output power from CDM is -9.5 dBm. Both amplified and un-amplified cases have been investigated experimentally. In the un-amplified case, we added a pair of attenuators (ATTs) in both the signal and the remote LO path, i.e., ATT1, to adjust the link losses of the two paths simultaneously. In the amplified case, we add one more pair of 21dBm EDFAs and ATTs in the link, as shown in Figure 3. 15(i). The ATT1s emulate the loss of

a 16-channel dense WDM multiplexer (MUX) and patch panel loss. Since the ATT1 loss changes the OSNR, an optical spectrum analyzer (OSA) was added to monitor the OSNR after EDFA. Another pair of ATTs after EDFAs were used to emulate the link loss (and dense WDM de-multiplexer (DEMUX) loss for the signal path) and also adjust the signal and LO ROP. A 1.5nm bandpass optical filter was placed before the receiver to remove the out-of-band amplified spontaneous emission noise. At Rx, a pair of polarization controllers (PCs) were used to recover the signal and remote LO state of polarizations (SOPs), respectively. The LO SOP is recovered to linear polarization by an automatic PC. The signal SOP is recovered by minimizing the X and Y polarization crosstalk, which can be achieved by monitoring the amplitude of low frequency (\sim tens of kHz) dither tones inserted in each tributary in both polarizations. An ICR with a bandwidth of 64GHz receives the signal and remote LO. The electrical signals were sampled by four 256 GS/s ADCs, and then processed by the offline DSP. We compare three DSP schemes as shown in Figure 3. 15(ii) and (iii), which are 4×2 butterfly MIMO equalizer (EQ) at 2 Sa/symbol (as an ideal reference), simplified 4×2 MIMO EQ at 1.5 Sa/symbol (the proposed scheme), and PAM4 EQ at 1 Sa/symbol (the simplest DSP). After resampling and synchronization, the MIMO EQ or the simplified MIMO EQ was applied, and followed by signal demodulation and BER counting. Since there is a slow time-varying I-Q phase rotation due to the path length mismatch between the signal and LO and/or ambient temperature and stress changes in the two fibers spools, a constant phase de-rotation (update period $\approx 3.33 \times 10^{-5}$ s) was applied before PAM4 EQ. This is due to that PAM4 EQ treats I and Q independently and cannot deal with such a phase rotation. Figure 3. 15(a) to (c) show the three least-mean-square-based DSP structures. The 4×2 MIMO EQ considers both the I&Q

crosstalk and X&Y crosstalk while the simplified 4×2 MIMO EQ considers only the I&Q crosstalk. The complexity of our proposed simplified 4×2 MIMO is similar to an I&Q correction scheme [89], although the difference resides in that our algorithm minimizes the error of complex QAM symbols while [89] minimizes the error of each PAM symbol. Both the MIMO and simplified MIMO use 128 EQ taps. The PAM4 EQ with 129 taps processes the four PAM4 data independently, which can only remove the ISI due to bandwidth limitation.

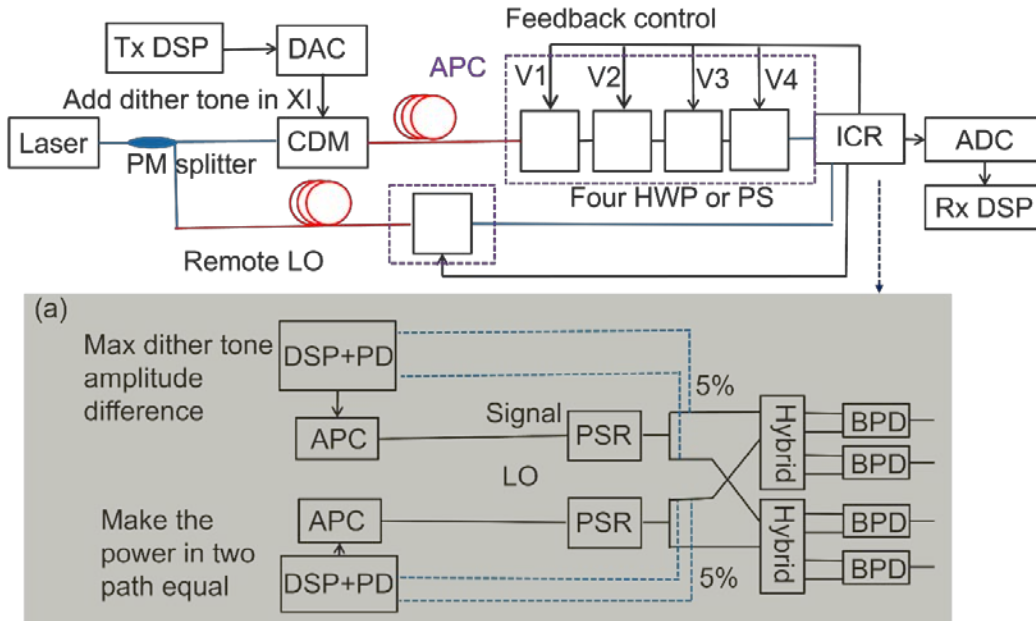


Figure 3. 16: Polarization tracking scheme in SHD coherent system.

We also preliminarily investigated the feasibility of reusing the low frequency (several kHz) dither tone in the automatic bias control scanning to do X and Y polarization de-multiplexing. The polarization tracking setup is shown in Figure 3. 16. For signal path polarization tracking, we use dither tone to do the signal polarization tracking. For instance, we insert the dither tone into XI at transmitter side. At the receiver side, as shown in Figure

3. 16, we tap out 5% optical signal after the polarization beam splitter inside ICR. Then the two 5% optical signals are detected by PD and processed by analog circuits to obtain the dither tone amplitudes in two paths. For signal polarization tracking, since we only insert dither tone into XI, we need to maximize the dither tone in X polarization path and minimize the dither tone in Y polarization path by doing feedback control to four or three half wave plates (HWP) or phase shifters. As for LO path polarization tracking, it is quite simple. The LO signal have been recovered by doing feedback control on two phase shifters or one HWP in and to make the LO power after the PBS having equal power.

We run a gradient descent algorithm to update the four voltage values applied to the four waveplates in automatic PC (APC). In this dissertation, the algorithm is implanted offline through python. Let the voltages applied to the APC in t-th iteration be $v^{(t)}$ and the voltages in (t+1)-th iteration be $v^{(t+1)}$. The gradient descent algorithm updates the control voltages to the APC as following:

$$v^{(t+1)} = v^{(t)} + \gamma \cdot \left(f(v^{(t)} + \mu \xi^{(t)}) - f(v^{(t)}) \right) \cdot \xi^{(t)}, \quad (40)$$

where $v^{(t)} = [V_1^{(t)} V_2^{(t)} V_3^{(t)} V_4^{(t)}]$, and $\xi^{(t)} \sim \mathcal{N}(0, I)$. γ is the step size and μ is the smoothing parameter. $f(x)$ stands for the dither tone amplitude difference value with the applied voltages equal to x . An adaptive step size (which decays with the increasing of iteration times) shows a fast convergence rate with ~ 10 iteration times.

3.3.2 Experimental and Simulation Results

First, we experimentally evaluate the link budget of the BtB unamplified link without fiber dispersion case as shown in Figure 3. 17. The result is with 0m path mismatch

and after applying a constant phase de-rotation to the whole frame (114688 symbols) in PAM4 EQ. The total link loss represents the loss of both the signal path and LO path. With respect to a low latency FEC BER threshold of 4×10^{-3} , PAM4 EQ can only support 3-dB link loss while 4×2 MIMO and simplified 4×2 MIMO can support 5.9 dB and 5.2 dB, respectively.

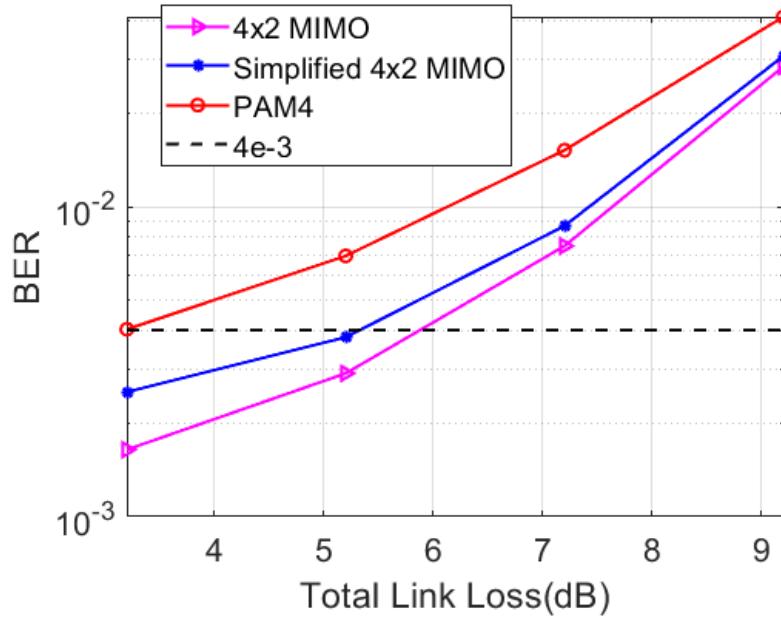


Figure 3. 17: Experimental results (link budget) of unamplified link. ©OSA 2022

Next, we investigate the amplified links with 0m path mismatch and constant phase de-rotation to the whole frame. Figure 3. 18(a) presents the BER performance versus ROP with different remote LO power into the ICR using 4×2 MIMO EQ. The optimal ROP for LO = 11 dBm is -5 dBm, whereas it changes to 0dBm for a reduced LO = 3 dBm. In the following experiments, we fix the received remote LO as 3 dBm and the signal ROP as -0.5 dBm to emulate a link loss of 5.5 dB ($=21-10 \times \log(16)-4-(-0.5)$) between the EDFA and DEMUX, which considers that the 21dBm EDFA output power is shared by 16 DWDM

channels, and a DEMUX loss of 4dB. The receiver sensitivity using a conventional LO and a 1:2 split ratio between LO and signal was also presented for comparison [112], which shows a similar performance to the SHD case of LO = 11dBm. Next the scheme's OSNR tolerance was found by tuning the ATT1 to 0 dB, 5dB, 9dB, and 15dB, respectively. As shown in Figure 3. 18(b), the PAM4 EQ gives the worst performance and exceeds the BER threshold when OSNR is smaller than 40 dB. We assume that the loss due to DWDM multiplexer and patch panel before EDFA is 5dB for both the signal and LO paths and obtain a fixed OSNR of 37.5 dB.

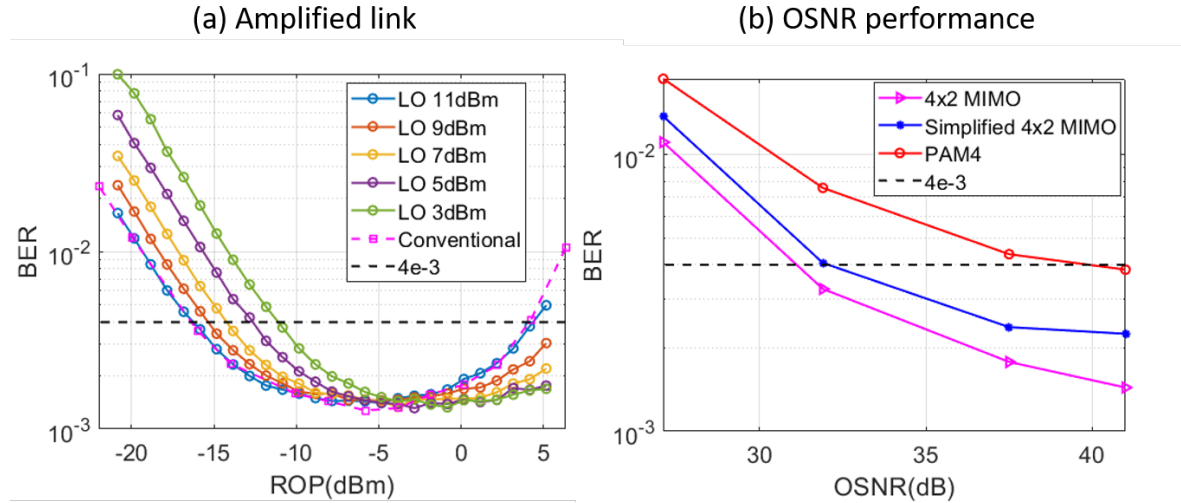


Figure 3. 18: (a) BER vs ROP and (b) BER versus OSNR (LO = 3dBm, signal ROP=-0.5dBm) of amplified link. ©OSA 2022

Figure 3. 19(a) presents the BER performance as a function of path length mismatch between the signal and LO. The 4×2 MIMO and modified 4×2 MIMO show very small BER degradation when path mismatch varies from 0 to 50m, i.e., both are not sensitive to the phase variation due to path mismatch. This is because these two EQs minimize I and Q

crosstalk, and therefore no additional phase de-rotation is needed. As for PAM4 EQ, a clear BER degradation can be observed for increased path length mismatch.

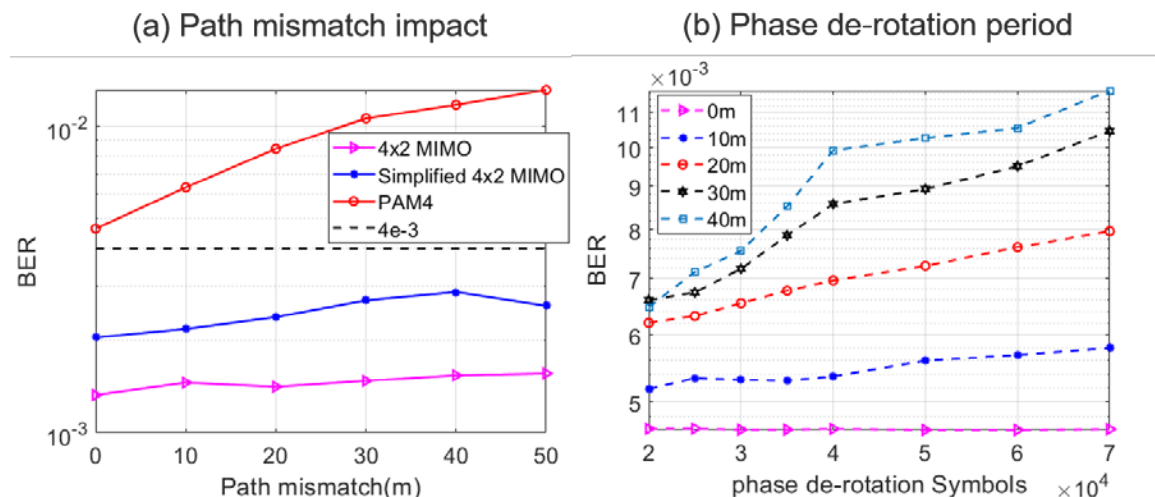


Figure 3. 19: BER vs path mismatch length and de-rotation period (OSNR = 37.5dB, LO=3dBm, signal ROP = -0.5dBm). ©OSA 2022

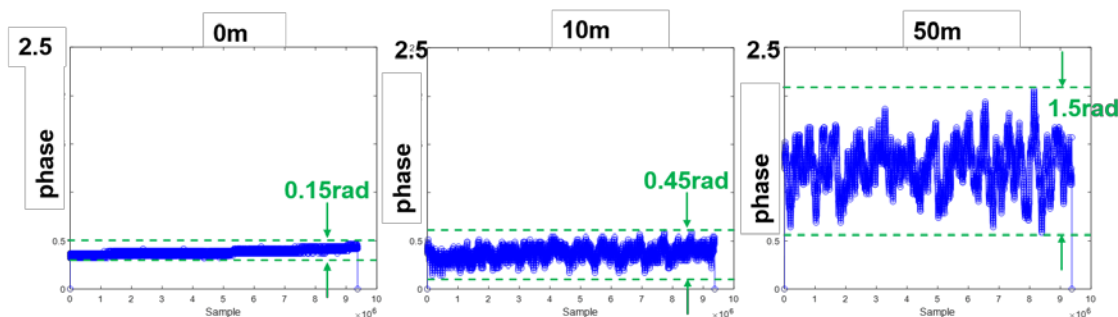


Figure 3. 20: Phase variation over time of the captured signal.

Figure 3. 19(b) shows how the BER changes with different phase de-rotation update period (the phase rotation angle varies with time and therefore the corresponding digital phase de-rotation must be constantly updated) in terms of number of symbols when PAM EQ is used. The BER of 0m path length mismatch is independent of the phase de-rotation update period while for 10 to 50m mismatch the BER increases when the phase de-rotation

update period increases. Figure 3. 20 shows the measured phase over time in the experiment and it validates Figure 3. 19(a). With the increasing of path mismatch length, the phase variation becomes much more significant.

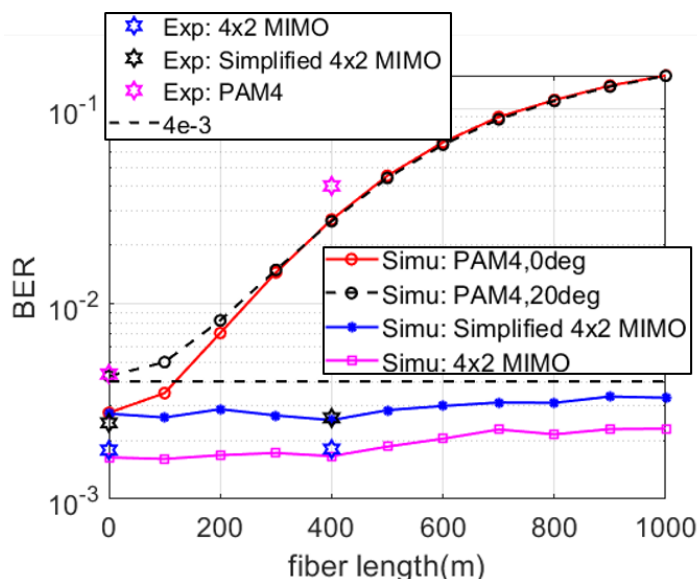


Figure 3. 21: BER versus fiber length with 0m path mismatch. ©OSA 2022

Figure 3. 21 shows the simulation (in solid lines) and experimental results (in star symbols) of BER versus fiber transmission length, and they match well. When the fiber length is 0m, the PAM4 EQ can have the same BER performance as that of the simplified MIMO, but two factors make PAM4 EQ's BER further degrade: (1) when the I-Q rotation angle is nonzero (e.g., 20-degree in Figure 3. 21), and (2) fiber chromatic dispersion. Factor (1) is caused by the finite ADC resolution (e.g., ENOB=5.5 in the simulation) which forbids a precise de-rotation angle of the I-Q constellation. We have verified by simulation that PAM4 EQ causes significant BER degradation when both receiver skew and I-Q phase rotation exist, while a higher ADC resolution (e.g., ENOB=13) can improve the performance. Factor (2), which can be modeled as a complex finite impulse response (FIR)

filter in time domain causes frequency-dependent I-Q mixing and ISI [113]. As shown in Figure 3. 21, modified 4×2 MIMO EQ can easily exhibit a BER lower than the 4×10^{-3} threshold up to 1km of standard single-mode fiber, while PAM4-based DSP cannot achieve a BER lower than the threshold considering a randomly rotated received I-Q constellation.

3.3.3 Summary

In a summary, through experiments and simulations, we have demonstrated a 120-Gbaud DP-16QAM self-homodyne coherent system with link budgets of 4.5 dB via a low-latency FEC and a simplified 4×2 MIMO DSP algorithm. The system architecture is more cost-efficient than the conventional coherent systems, and thus it is promising for future intra-data center applications. In the next section, we will introduce efficient algorithm design with improved system performance.

3.4 Data-Efficient Nonlinearity Correction Algorithm

As indicated in chapter 1, the conventional VNLE needs a large training overhead and will decrease the transmission efficiency in the case of burst frame, multi-user links, or dynamic channel conditions, where the taps need to be updated within several μ s or ms intervals. Therefore, to further improve the VNLE performance and reduce the requirement on training data size, we investigate both the semi-supervised and supervised VNLEs and perform an experimental validation in a fiber-FSO converged transport system [78, 79].

3.4.1 Principle of Operation

To mitigate the nonlinearities in the system, the full VNLE fits the nonlinear transfer curve can be represented by the following class of polynomials [69]:

$$\begin{aligned}
z(n) = f_{full}^{(n)}(h, s) = & \sum_{k=-(K-1)/2}^{(K-1)/2} h_k \cdot s(n-k) \\
& + \sum_{p=2}^{N_p} \sum_{i_1=-(L-1)/2}^{(L-1)/2} \dots \sum_{i_p=-(L-1)/2}^{(L-1)/2} h_{i_1, \dots, i_p} \cdot \prod_{i_p} s(n-i_p),
\end{aligned} \tag{41}$$

where s is the input data, z is the output data, and h are the tap coefficients. N_p determines the polynomial orders in the VNLE. High N_p can fit the nonlinear model better but requires higher computational complexity. For a given sequence s , the notation $f(h, s)$ stands for the output sequence of VNLE parametrized with h applied to s . We use $f^{(i)}(h, s)$ to denote the i -th element of such sequence. Also, a simplified VNLE (sVNLE) is widely deployed to reduce the computational complexity. It only keeps the power terms and is presented as [67, 70]:

$$\begin{aligned}
z(n) = f^{(n)}(h, s) \\
= \sum_{k=-(K-1)/2}^{(K-1)/2} h_k \cdot s(n-k) + \sum_{p=2}^{N_p} \sum_{i=-(L-1)/2}^{(L-1)/2} h_{i,p} \cdot s^p(n-i).
\end{aligned} \tag{42}$$

With known input samples and known labels, the estimator chooses the tap coefficients that minimize the least square error between the estimated output samples and the labels. This method is the conventional VNLE using OLS, which utilizes the labeled data as the training data (i.e., pilot symbols). In the following analysis, our methods generalize it by using the unlabeled data and sparsity-inducing regularization.

Let X denote the labels at the transmitter side and let \hat{X} denote the received labeled data. The objective function for the supervised mode with Lasso consists of two terms, the

mean-squared error on the training symbols $\|f(h, \hat{X}) - X\|_2^2$, and an ℓ_1 regularization term.

The supervised method objective is thus formulated as a convex optimization program:

$$h = \arg \min \left(\|f(h, \hat{X}) - X\|_2^2 + \lambda \cdot \|h\|_1 \right). \quad (43)$$

When $\lambda = 0$, it degenerates into the supervised VNLE using OLS estimator. The added regularization term can exploit the sparsity structure in the optimal coefficients. Even if the ambient dimension of data vectors is much larger than the number of observations, the Lasso estimator can still recover the optimal coefficients, provided that it is (approximately) sparse [114, 115]. Therefore, it can reduce the requirement on training data size. In addition, the lasso estimator automatically performs model selection: when superfluous high-order coefficients are included, its weight will typically become zero. The regularization parameter λ can be chosen through cross validation or the theoretical formula [114]. In this paper, we choose λ using 10-fold cross validation. Precisely, the algorithm partitions the training data into 10 folds, and use 10% of the training data as a validation dataset to choose λ . This methodology is known to be able to select the optimal value of λ in Lasso [116]. For each value of λ , we perform cross-validation separately, yielding the optimal λ in each case. The Lasso program can be solved via alternating direction method of multipliers (ADMM). The OLS problems are solved via method of steepest descent in our paper. In each iteration of ADMM for Lasso, the complexity is a matrix-vector product, which is the same as the per-iteration cost of steepest descent for OLS. The ADMM algorithm used in this paper belongs to the class of gradient-based methods. When applied to Lasso with high-dimensional problems, the convergence rates of such methods typically

depend on the Restricted Strong Convexity condition of the problem [117]. This is also comparable to steepest descent, for which the convergence rates of gradient-based methods depend on the condition number of the data matrix [118].

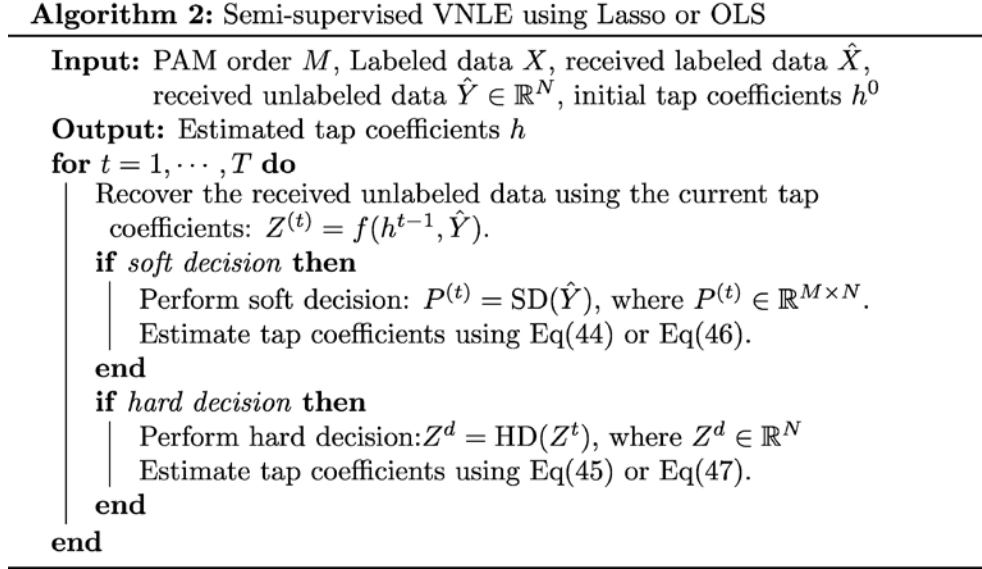


Figure 3. 22: Semi-supervised VNLE with Lasso using SD and HD. ©IEEE 2021

Moreover, we propose to use semi-supervised methods using self-training (also known as pseudo-label) in polynomial regression [80, 81]. It exploits the unlabeled data and further reduces the amount of labeled data needed. This type of methods may get stuck in some cases as the decision error made by the algorithm can reinforce itself, instead of being corrected. Other approaches, such as MixMatch [119] and temporal ensembling [120] may be used to improve the semi-supervised algorithms, albeit costing higher computational complexity. In this work we will focus on the simple self-training method. We use an iterative algorithm, which alternates between estimating the labels and estimating the coefficients. The algorithm takes the labeled data X , received labeled data \hat{X} and received unlabeled data \hat{Y} as inputs. We will consider and investigate both the HD

and SD in the decision step. HD is performing maximum likelihood detection on each symbol independently (i.e., based on minimum Euclidean distance). The SD probabilities are computed through the posterior distribution of the input labels given the observed value, where the noise is assumed to be Gaussian. For each iteration, the following computations are performed sequentially as shown in Figure 3. 22. First, the algorithm recovers the received data using the current Volterra series. Second, perform SD or HD. The SD output is a probability matrix that denotes the probability of the symbols belong to each alphabet. The HD output is the symbol sequence after decision. Let N be the number of the unlabeled symbols and M be the alphabet size (e.g., $M = 8$ for PAM 8), the VNLE tap coefficients at the t -th iteration using SD and HD are estimated as

$$h^t = \arg \min_h \left(\sum_{i=1}^N \sum_{\beta=1}^M P^{(t)}(\beta, i) \cdot \left(f^{(i)}(h, \hat{Y}) - \beta \right)^2 + \alpha \cdot \left\| f(h, \hat{X}) - \mathbf{X} \right\|_2^2 + \lambda \|h\|_1 \right), \quad (44)$$

and

$$h^t = \arg \min_h \left(\left\| f(h, \hat{Y}) - Z^d \right\|_2^2 + \alpha \cdot \left\| f(h, \hat{X}) - \mathbf{X} \right\|_2^2 + \lambda \|h\|_1 \right), \quad (45)$$

respectively. In equation 44, $P^{(t)}(\beta, i)$ is the soft-decision probability of the i -th unlabeled symbol being symbol β in the alphabet at the t -th iteration. In equation 45, $Z^d \in \mathbb{R}^N$ is the symbol sequence generated by HD. Compared to equation 43, the objective function equation 44 and 45 contain an additional term, which is the weighted mean-squared error on the unlabeled data. This additional term aims at minimizing the difference between the

signals after and before the decision. We add a coefficient α to tune the weight between the labeled and unlabeled data. When $\alpha=0$, the estimator becomes un-supervised mode, which only utilizes the unlabeled data and does not need training symbol or introduce overhead. When $\lambda = 0$, they degenerate into semi-supervised VNLEs using OLS, which are presented as:

$$h^t = \arg \min_h \left(\sum_{i=1}^N \sum_{\beta=1}^M P^{(i)}(\beta, i) \cdot \left(f^{(i)}(h, \hat{Y}) - \beta \right)^2 + \alpha \cdot \left\| f(h, \hat{X}) - X \right\|_2^2 \right), \quad (46)$$

and

$$h^t = \arg \min_h \left(\left\| f(h, \hat{Y}) - Z^d \right\|_2^2 + \alpha \cdot \left\| f(h, \hat{X}) - X \right\|_2^2 \right). \quad (47)$$

The semi-supervised sVNE involves additional unlabeled data in multiplications and needs several times of decision-direct iterations, which shows higher complexity than supervised sVNLE.

Proper initial tap coefficients reduce the iteration times of the semi-supervised VNLE using OLS and help to avoid bad local minima. We tried two tap initializations. One approach is a simple initialization that has been used for linear equalizer: the initial tap coefficients in equation 42 are set to $h_0=1$, while the rest of the coefficients are set to zero. The other is called supervised initialization, where the output of supervised VNLE using Lasso serves as the initial tap. The initial BER of supervise initialization is the BER of supervised VNLE using Lasso. The algorithm terminates under certain stopping criteria. In the following experimental evaluation, the algorithm terminates when the Euclidean

distances between tap coefficients obtained by 3 consecutive iterates are less than a certain threshold ε (i.e., 0.001) or the iteration times exceed 50. We use normalized values of symbol energy in our algorithms.

3.4.2 Experiment Setup

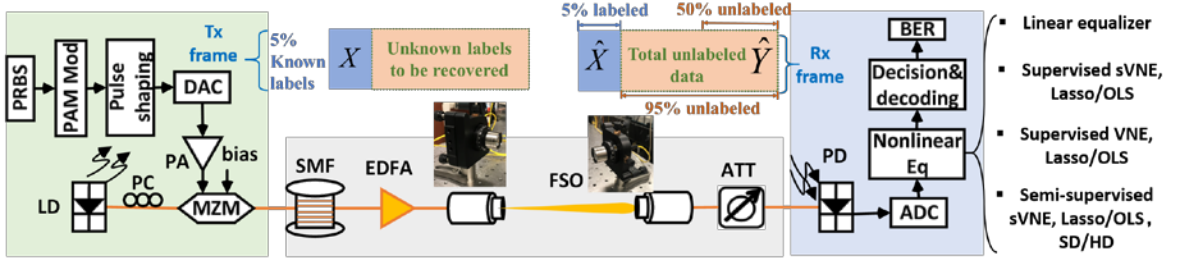


Figure 3. 23: Experiment setup of the fiber-FSO converged link. ©IEEE 2021

Figure 3. 23 shows the experimental setup of the fiber-FSO link. At the transmitter side, the PRBS sequence is mapped to PAM signal and then pulse shaped using a root-raised-cosine filter with roll off factor equals to 0.01 in offline DSP. The symbol number is 10^5 , which includes both the labeled data and the unlabeled data. The offline generated signal is loaded into DAC and then amplified by a 25-dB PA. The DAC resolution is 8 bits, and its bandwidth is 25 GHz. The transmitter LD operates at the wavelength of 1550 nm and its power is set to be 12 dBm. The electrical signal is then modulated by a 40-GHz MZ Modulator with the V_π of 5.2 V. The optical signal propagates through a 5-km standard SMF with 1-dB loss and then amplified by an EDFA. The EDFA output power is 18.32 dBm. A pair of collimators transmit the signal through a 2-m free-space optical link with 5.32-dB loss. At the receiver side, an ATT adjusts the received signal power for conducting the sensitivity evaluation. A 50-GHz PIN PD detects the optical signal, which is followed by an 80-GSa/s ADC and offline DSP. The ADC resolution is 8 bits, and its bandwidth is

25 GHz. The VNLEs recover the received signal and then signal decision and decoding are performed to count the BER for performance evaluation. Several equalizer schemes are investigated in the next section. Linear equalizer only keeps the linear (first order) ISI. Thus, it does not compensate for the nonlinear impairments that result in higher order polynomial terms. For the nonlinear equalizer, the sVNLE will be investigated with supervised and semi-supervised methods using Lasso or OLS. The insets of Figure 3. 23 illustrate an instance of the transmitter data frame and receiver data frame with 5% labeled data. We will sweep labeled data ratio and unlabeled data ratio in the next section. The labeled data ratio is pilot symbol ratio. As for unlabeled data ratio, it means the ratio of the received data that need to be recovered in the semi-supervised methods. For instance, in our experiment, there are 10^5 symbols in each frame. The 5% labeled data means $0.05 \times 10^5 = 5 \times 10^3$ symbols. The 20% unlabeled data means $0.2 \times 10^5 = 2 \times 10^4$ symbols. For BER evaluation, we use 9×10^4 unlabeled symbols considering the highest portion of labeled data is 10% in the investigation.

3.4.3 Experimental Results

First, we sweep and optimize the semi-supervised VNLE settings including the tap initialization, unlabeled data ratio, memory lengths, as well as the weight (α in equation 44 to 47) between the labeled data and unlabeled data. The peak-to-peak voltage (V_{pp}) is set as 4.2 V for 50-Gbaud PAM4 and 35-Gbaud PAM8. The ROP is 6.6 dBm. The memory lengths are set as $K = 128$, $L = 15$, $N_p = 5$. Figure 3. 24 presents the BER versus α under two different initialization methods for semi-supervised VNLEs. The optimal BER and α can be obtained in the “dip” of each curve. Larger α gives more weight to the labeled data. Using simple initialization, the number of iterations required by the algorithm using OLS

(from 15 to 35 typically) is much higher than that using Lasso (< 10). On the other hand, when combined with Lasso, it is harder for the semi-supervised algorithm to converge to a good local minimum. The Lasso program needs enough amount of weight on true labels, so that the support of the sparse vector can be approximately recovered in the initial iterations.

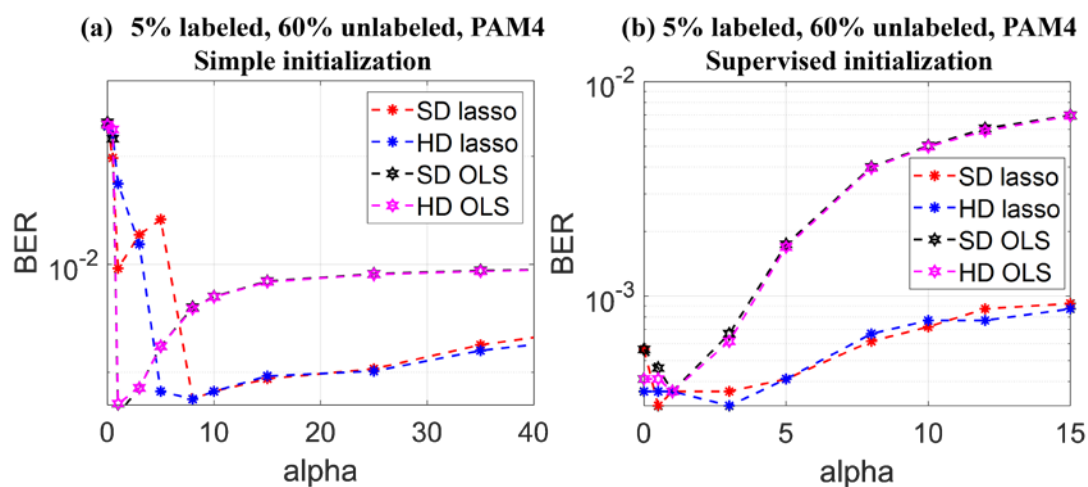


Figure 3. 24: BER versus weight α for 50-Gbaud PAM4 (5% labeled data and 60% unlabeled data). ©IEEE 2021

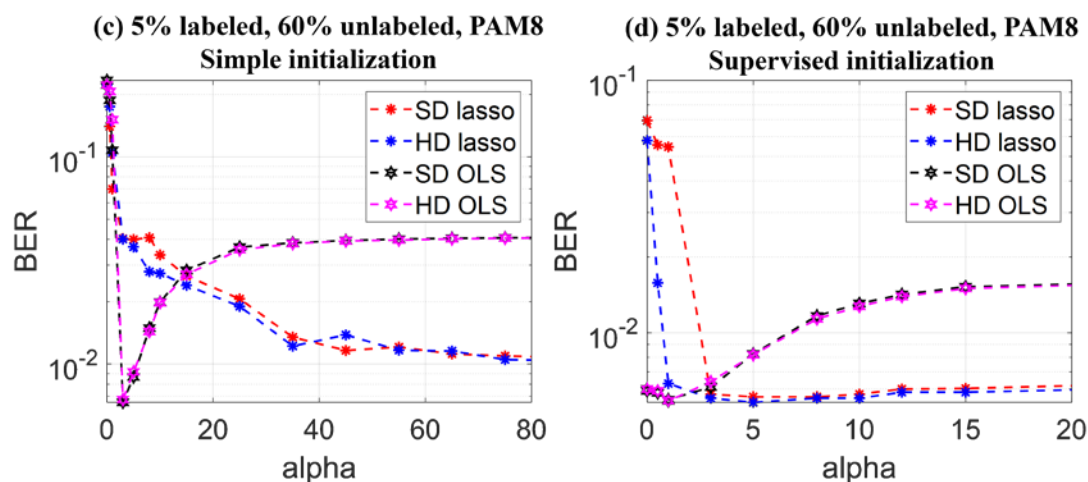


Figure 3. 25: BER versus weight α for 35-Gbaud PAM8 (5% labeled data and 60% unlabeled data). ©IEEE 2021

This situation is aggravated with PAM8, where the structure of non-linearity is more complicated. As shown in Figure 3. 25(a), the optimal BER of OLS is better than Lasso in PAM8 signal under simple initialization. Supervised initialization utilizes the output of supervised VNLE using Lasso and help the algorithms to get rid of the bad local optimal. As shown in Figure 3. 24(b) and Figure 3. 25(b), Lasso shows better or comparable performance than OLS, and the optimal BER of OLS is also improved compared with the simple initialization. The optimal alpha of supervised initialization (<5) becomes smaller than that under simple initialization, which means the algorithm puts more weight on the unlabeled data due to better initial decision errors. Moreover, both semi-supervised methods require smaller iterations (3~7) to converge, under the supervised initialization. In the following analysis, we will use the supervised initialization to get better results.

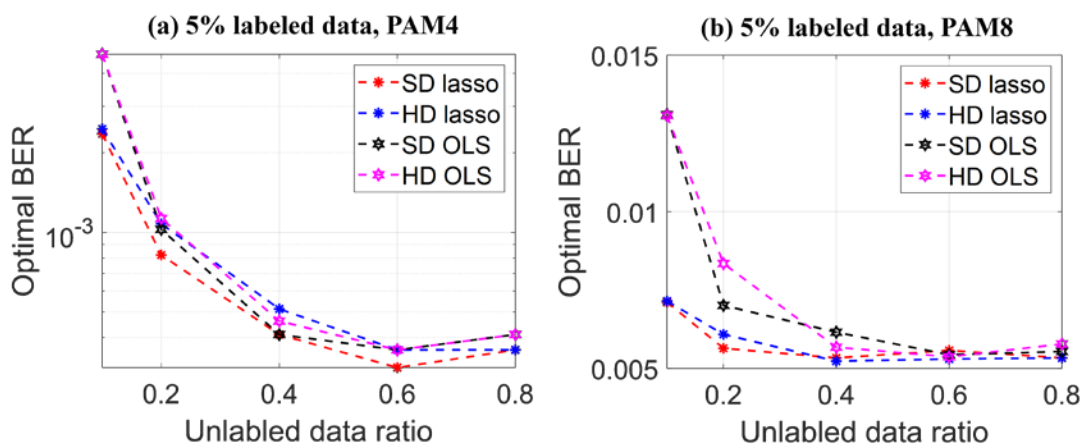


Figure 3. 26: Optimal BER versus unlabeled data ratio with 5% labeled data for 50-Gbaud PAM4 and 35-Gbaud PAM8. ©IEEE 2021

Then by sweeping the weight α , we obtain the optimal α and BER for different unlabeled data ratios as shown in Figure 3. 26. Figure 3. 26(a) and (b) show the optimal BER obtained from the optimal α with different unlabeled data ratios and 5% labeled data

for PAM4 and PAM8, respectively. Overall, the performances improve with the increasing of the unlabeled data ratio. When the unlabeled data ratio is small (i.e., $\leq 20\%$), Lasso estimator shows better BER performance than OLS. When the unlabeled data ratio is large (i.e., $>20\%$), the performances of OLS are similar to that of Lasso.

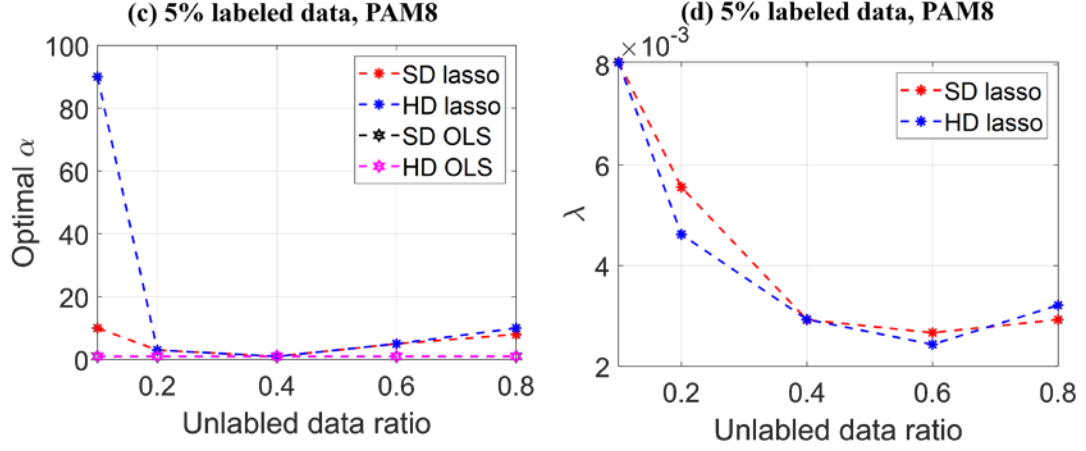


Figure 3. 27: Optimal α and λ versus unlabeled data ratio with 5% labeled data for PAM8. ©IEEE 2021

Figure 3. 27(a) present the optimal α with respect to different unlabeled data ratios using 5% labeled data for 35-Gbaud PAM8. When the unlabeled data ratio is small (i.e., 10%), Lasso uses large value of α (i.e., 90) to put more weight on labeled data due to insufficient unlabeled data. When the unlabeled data ratio is larger than 10%, the optimal α becomes smaller (<10). The optimal weights of OLS remain unchanged and the optimal α is 1. As mentioned in previous section, we use 10-fold cross validation to select the optimal value of λ in lasso. Figure 3. 27(b) shows the optimal λ with different unlabeled data ratios. The value of λ decreases as the unlabeled data size increases. It is due to that the increasing of the observation number can relax the regularization term.

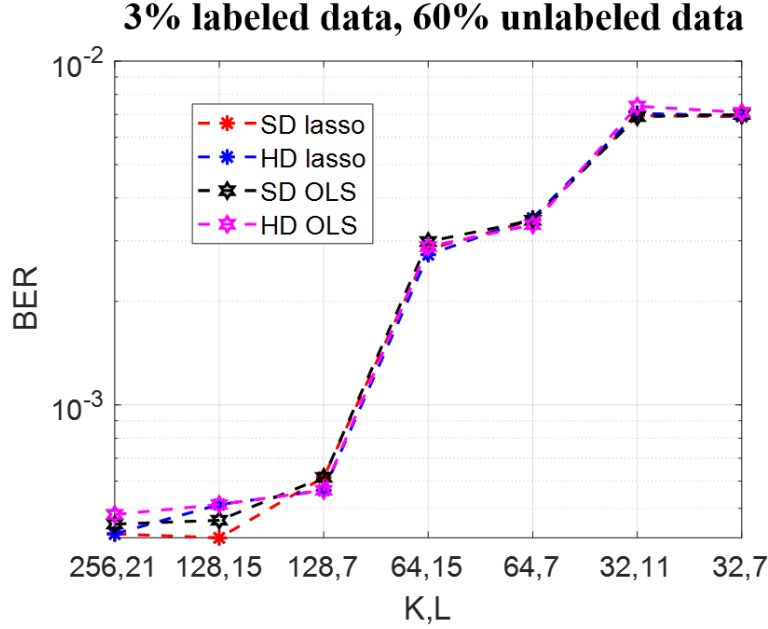


Figure 3. 28: BER of 50-Gbaud PAM4 versus memory lengths using 3% labeled data and 60% unlabeled data. ©IEEE 2021

After sweeping the weight and unlabeled data ratios, we sweep different memory lengths as presented in Figure 3. 28. The unlabeled data ratio is 60% and the labeled data ratio is 3%. With the decreasing of the linear and power term memory lengths, the BER becomes worse. Note that using even larger memory lengths (i.e., $K = 256$, $L = 21$) cannot further improve the system performance but brings more complexity. Thus, the best BER value is attained by setting $K = 128$ and $L = 15$. As shown in Figure 3. 26 to Figure 3. 28, the four semi-supervised algorithms show comparable or similar BER performances in most cases. In some cases, the semi-supervised sVNLE using SD and Lasso shows slightly better performances over other semi-supervised algorithms. Intuitively, this is because SD makes use of the log-likelihood information, instead of only the binary decision. Also, Lasso can reduce the number of tap coefficients and implementation complexity in semi-supervised methods. Thus, in the following analysis, for simplification, we will use SD and

Lasso for semi-supervised sVNLE. In practice, HD could be considered due to its lower computational complexity and similar BER performances to SD in most cases. Also, we set $K = 128$, $L = 15$ to have the best BER performance according to the previous results. The unlabeled data ratio is set as 60%.

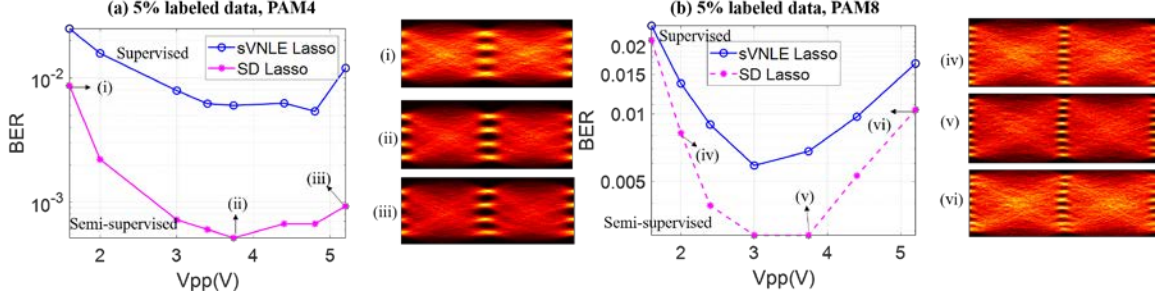


Figure 3. 29: BER versus V_{pp} and signal eye diagrams using 5% labeled data.
©IEEE 2020

Figure 3. 29 compares the performances of supervised VNLE using Lasso and semi-supervised VNLE using Lasso and SD with different peak-to-peak voltages (V_{pp}) into the modulator. The insets show the eye diagrams after nonlinear compensation under different V_{pp} . Figure 3. 29(a) and (b) present the results of 50-Gbaud PAM4 signal and 35-Gbaud PAM8, respectively. With 60% unlabeled data, the semi-supervised VNLE using Lasso and SD can improve the BER value by one order of magnitude. As shown in the eye diagrams, lower V_{pp} yields a lower signal to noise ratio (SNR) and lower modulation extinction ratio. Thus, increasing the V_{pp} can improve the signal SNR and BER. However, when V_{pp} exceeds the linear operation region, nonlinearity degrades the signal quality. When V_{pp} is approaching V_{π} and the nonlinearity become significantly high, the signal performance would still degrade even if the VNLE is used. One can observe that in insets (vi), the eye diagram still gets compressed after nonlinear correction. Based on these

results, in the following analysis, we choose the V_{pp} of the PAM4 and PAM8 as 4.2V and 3.75V, respectively.

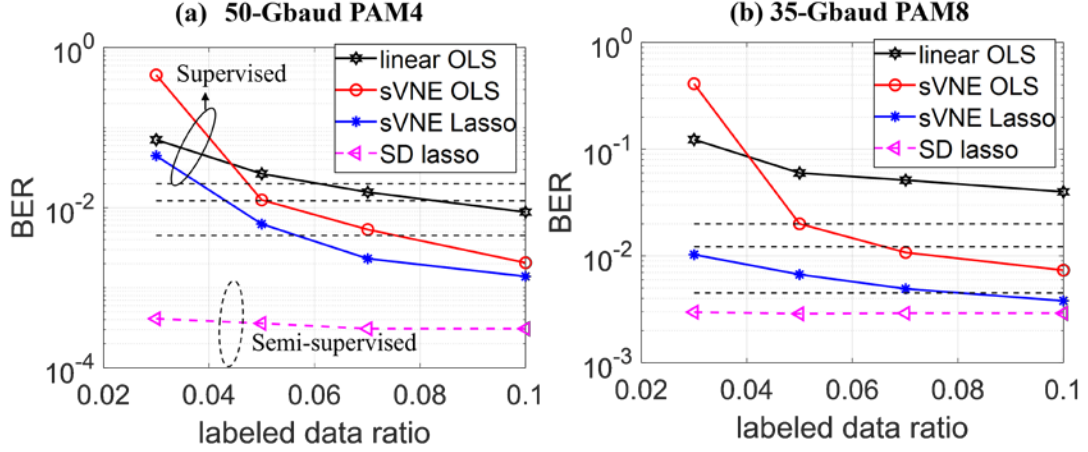


Figure 3. 30: BER versus labeled data ratio using 60% unlabeled data. ©IEEE 2020

Figure 3. 30(a) and (b) show the BER versus labeled data ratios using different VNLEs for both the 50-Gbaud PAM4 ($V_{pp}=4.2V$) and 35-Gbaud PAM8 ($V_{pp}=3.75V$). The ROP is 6.6 dBm. Three forward error correction codes (FEC) are considered and the corresponding pre-FEC BER thresholds are plotted as reference lines (dashed, horizontal). The thresholds are 4.5×10^{-3} , 1.22×10^{-2} and 2×10^{-2} , which are staircase FEC with 6.69% OH, concatenated FEC (cFEC) with 14.8% OH, openFEC (oFEC) with 15.3% OH, respectively. In both figures, the supervised sVNLE outperforms supervised sVNLE using OLS for the unlabeled data ratio ranging from 3% to 10%. The semi-supervised VNLE using SD and lasso can always achieve a lower BER value than the 4.5×10^{-3} threshold and show a relative flat curve when the labeled data ratio varies. Thanks to the unlabeled data, the semi-supervised VNLE show little degradation when the labeled data ratio is decreased. Moreover, for 35-Gbaud PAM8, linear equalizer fails to achieve the 2×10^{-2} FEC threshold

while supervised sVNLE using OLS fails to attain 4.5×10^{-3} threshold. Moreover, at the 2×10^{-2} threshold, the supervised sVNLE using lasso needs 3% labeled data while OLS needs 7% labeled data.

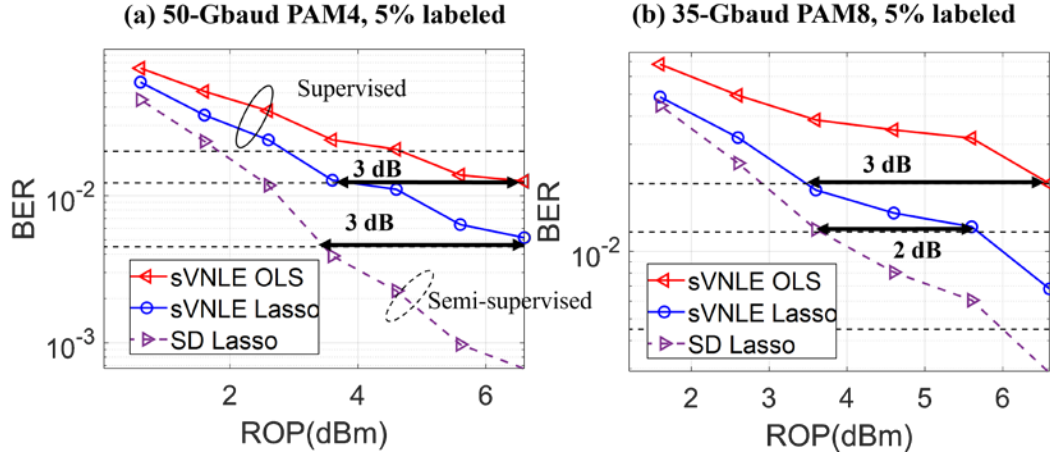


Figure 3. 31: BER versus ROP using 5% labeled data in FSO transmission. ©IEEE 2020

Figure 3. 31(a) and (b) show the sensitivity performance of 50-Gbaud PAM4 and 35-Gbaud PAM8 using 5% labeled data in FSO transmission. At the threshold of 4.5×10^{-3} in PAM4 transmission, the semi-supervised sVNLE using SD and Lasso demonstrates 3-dB sensitivity improvement in comparing with supervised sVNLE using Lasso. The supervised sVNLE using OLS fails to achieve the 4.5×10^{-3} threshold. At the 1.22×10^{-2} threshold in PAM4 transmission, the semi-supervised and supervised VNLE using lasso show 4-dB and 3-dB ROP improvement when compared with the supervised sVNLE using OLS, respectively. As for PAM8, the semi-supervised algorithm shows 2-dB gain over the supervised sVNLE using lasso at the 1.22×10^{-2} threshold. The supervised sVNLE using lasso achieves 3-dB sensitivity gain when compared with supervised sVNLE using OLS at the 2×10^{-2} FEC threshold. Moreover, the tap number reduction percentage was analyzed in

both cases to show the benefit of tap coefficients reduction. In 50-Gbaud PAM4 transmission, the supervised sVNLE using Lasso achieves 15% tap coefficients reduction while the semi-supervised sVNLE achieves 9% reduction. As for 35-Gbaud PAM8 transmission, the supervised sVNLE using Lasso and the semi-supervised sVNLE achieves 47% and 36% tap number reduction, respectively.

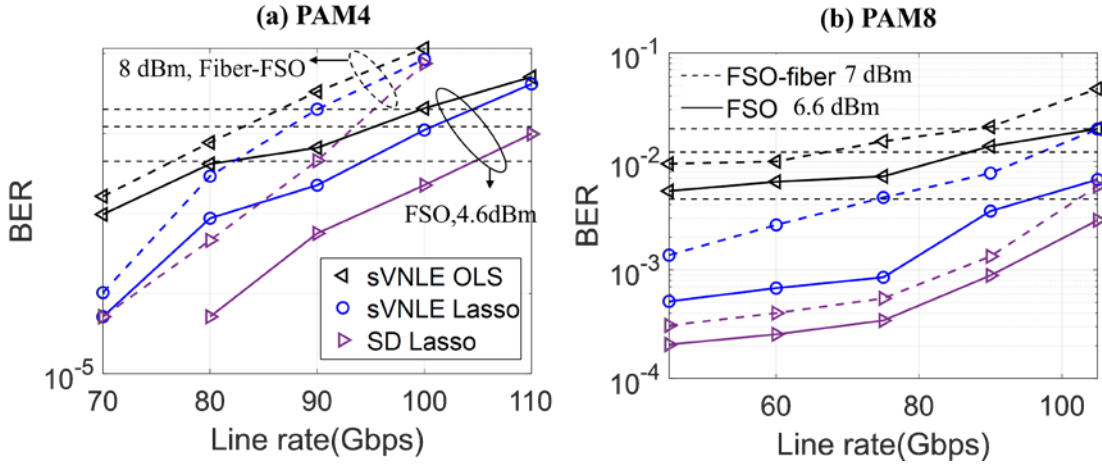


Figure 3. 32: BER versus line rates using 5% labeled data. ©IEEE 2020

Figure 3. 32(a) and (b) show the BER with different line rates using PAM4 and PAM8 with FSO and FSO plus fiber, respectively. Figure 3. 32(a) shows the results of PAM4 using 5% labeled data, where the ROP of FSO is 4.6 dBm and ROP of FSO-fiber is 8 dBm. In FSO transmission, the semi-supervised method can always support BER lower than the 2×10^{-2} or 1.22×10^{-2} threshold while other methods cannot support 110 Gbps at all three thresholds. In FSO-fiber transmission, semi-supervised sVNLEs can achieve 4.5×10^{-3} threshold at 90 Gbps while the supervised sVNLE using Lasso and OLS are at 80 Gbps and 75 Gbps. Figure 3. 32(b) shares the same legend as Figure 3. 32(a), which shows the PAM8 results. At the 4.5×10^{-3} threshold, the supervised sVNLE using OLS can only support 45 Gbps in FSO transmission while the supervised sVNLE using Lasso and semi-

supervised sVNLEs can support 95 Gbps and >110 Gbps, respectively. Thus, the proposed methods show a line rate improvement higher than 100% in this case. In FSO-fiber transmission (ROP = 7dBm), sVNLE using OLS can support 65 Gbps at the 2×10^{-2} threshold but fails to achieve 1.22×10^{-2} . The sVNLE using Lasso and semi-supervised sVNLEs can attain the 1.22×10^{-2} threshold at 75 Gbps and 100 Gbps.

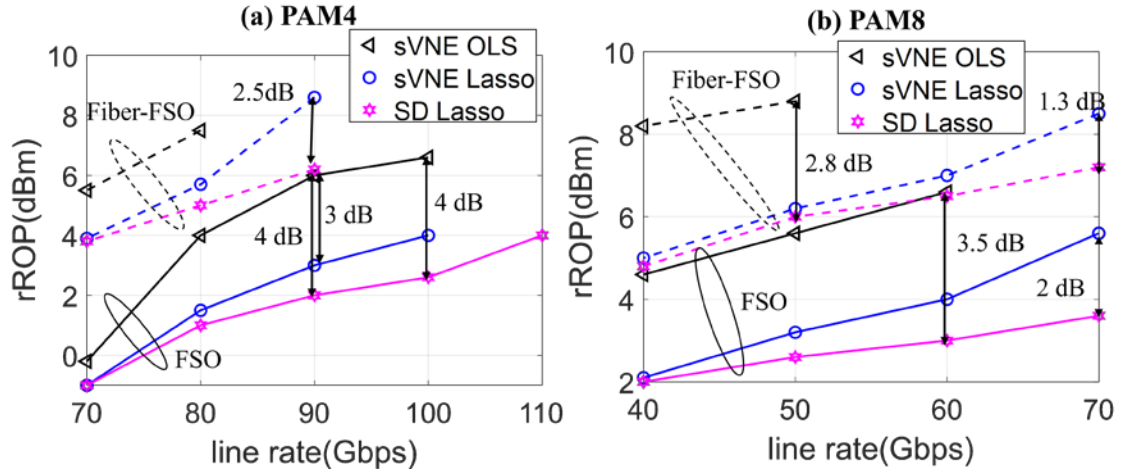


Figure 3. 33: Required ROP versus line rates using 5% labeled data at the 1.22×10^{-2} threshold. ©IEEE 2020

Figure 3. 33(a) and (b) present the required ROP (rROP) at 1.22×10^{-2} thresholds of PAM4 and PAM8 using 5% labeled data, respectively. Higher gains could be achieved using Lasso or semi-supervised method when the line rate is higher. This is due to that the system has fixed swing so the SNR will be lower when the line rate is higher. Lower SNR requires larger amount of data, rendering the data size in our experiments insufficient. In such case, the regularization effect by Lasso and semi-supervised methods can alleviate this issue and achieve better performance. In Figure 3. 33(a), semi-supervised sVNLE show up to 4-dB and 3-dB sensitivity gain over the supervised sVNLE using OLS and the supervised sVNLE using Lasso, respectively. In Figure 3. 33(b), the supervised sVNLE

using Lasso is better than the conventional method by up to 2.8 dB. Also, the semi-supervised sVNLE shows up to 3.5-dB and 2-dB ROP gain over the supervised sVNLE using OLS and using Lasso.

3.4.4 *Summary*

We leveraged and investigated semi-supervised methods and Lasso to VNLE, which can mitigate the nonlinearity in broadband communication systems. Experimental results in a fiber-FSO link validate that Lasso can reduce the required pilot symbol number by exploiting the sparsity of the tap coefficients. Among the supervised VNLEs, Lasso yields better performance. Experimental results showed that the supervised sVNLE using Lasso outperforms the supervised sVNLE using OLS by up to 3-dB sensitivity gain. Moreover, the semi-supervised VNLE further improved the BER performance, while maintaining the minimum frame overhead. With sufficient unlabeled data, our experimental results have shown that, in both the 50-Gbaud PAM4 transmission (at the 1.22×10^{-2} threshold) and 35-Gbaud PAM8 transmission (at the 2×10^{-2} threshold), the semi-supervised sVNLE using SD and Lasso demonstrates ROP gains up to 4-dB and 3-dB over the supervised sVNLE using OLS and the supervised sVNLE using Lasso, respectively.

CHAPTER 4. RELIABLE FIBER-WIRELESS SYSTEM DESIGN

4.1 Integrated mmW and FSO A-RoF Link

In this chapter we propose and experimentally demonstrate a CMC technique that enables the implementation of an ultra-reliable, CMC-based, integrated mmW/FSO A-RoF system [84]. For the access link with a few kilometers or less where fibers are not available, we employ the proposed A-RoF-based mmW/FSO system to enhance its robustness to different weather conditions and interferences.

4.1.1 Principle of Operation

The CMC system consists of two parts: coordinated mapping at the transmitter side and coordinated combining at the receiver side. Coordinated combining is the inverse process of coordinated mapping. These two parts optimize the system's reliability and performance.

At the transmitter side, the CMC mapper is specially designed for OFDM modulation formats widely adopted by 4G, 5G new radio, and Wi-Fi standards. As an example, assume that the number of the original data blocks in the time domain and frequency domain is N_t and N_f , respectively. The time_duration of the original signal is T and frequency spacing of each block is Δf . When the mapping ratio is M , the mapper expands the number of blocks in the time domain to $M \times N_t$ and the frequency spacing of the block to $M \times \Delta f$.

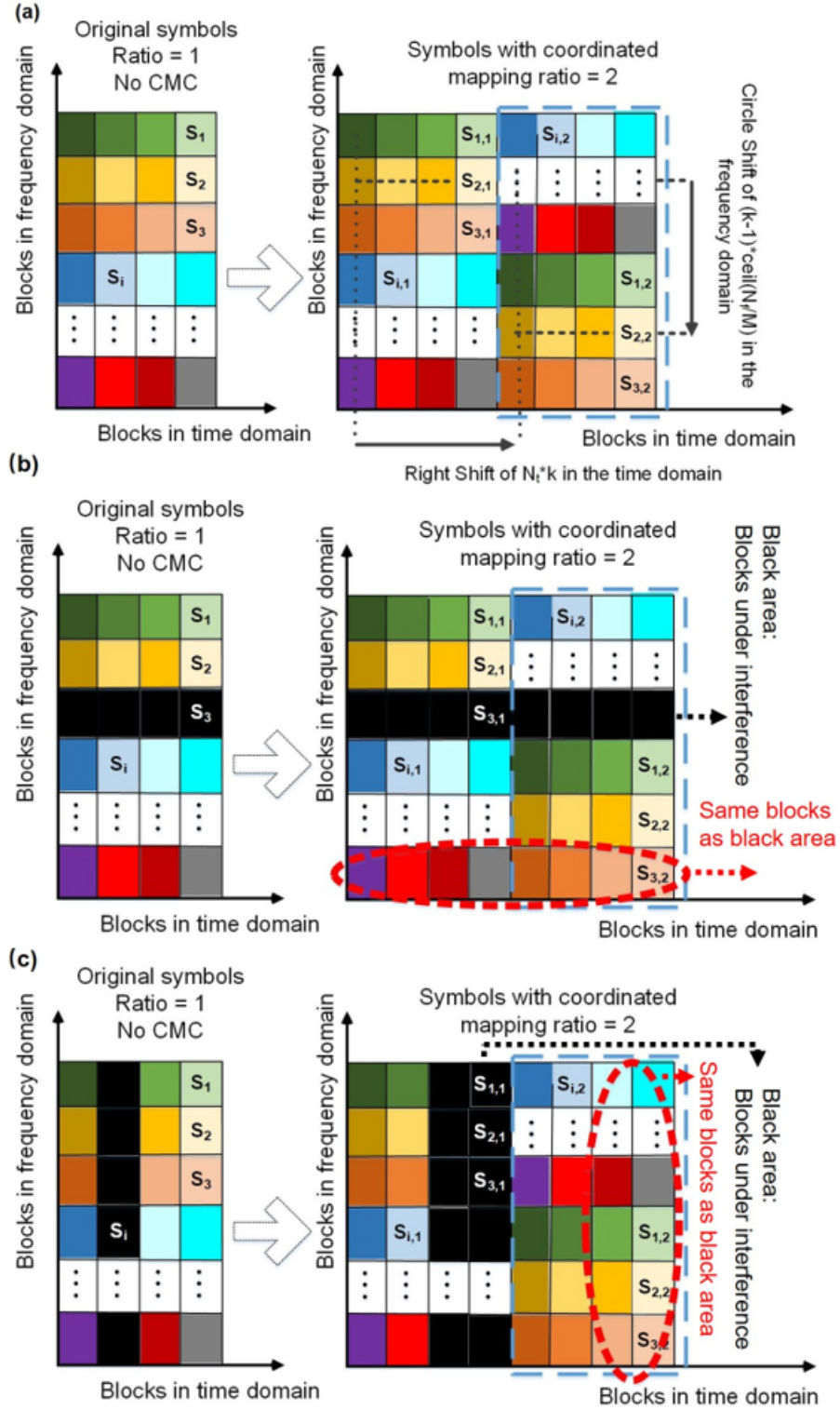


Figure 4. 1: Working principles of the coordinated mapping. ©IEEE 2018

The i_{th} block $S_i (1 \leq i \leq N_t \times N_f)$ of the original signal is $\Delta f_i = \Delta f$ wide in frequency and $\Delta t_i = \frac{T}{N_t}$ long in time. It is repeated and mapped to M blocks through the following equations:

$$S_{i,k} (\Delta t_{i,k}, \Delta f_{i,k}) = S_i (\Delta t_i, \Delta f_i), \quad (48)$$

$$\Delta t_{i,k} = \frac{\Delta t_i}{M} + (k-1) \times \frac{T}{M}, \quad (49)$$

$$\Delta f_{i,k} = \left[\left(\Delta f_i - \text{int} \left(\frac{N_f}{M} \right) \times (k-1) \times \Delta f \right) \bmod (N_f \times \Delta f) \right] \times M, \quad (50)$$

where $S_{i,k}$ is the same as $S_i (1 \leq k \leq M)$ as in equation 48. $\Delta t_{i,k}$ and $\Delta f_{i,k}$ are time duration and frequency increment of the k_{th} repeated block, respectively. In equation 50, the *int* operator rounds off to the nearest integer and the *mod* operator returns the modulus of a number. The mapping procedure is determined by mapping ratio M and the number of data blocks in the frequency and the time domains.

Figure 4. 1 (a) shows one example of mapping procedure with mapping ratio $M = 2$. Same color blocks represent same information content. Equation 49 shifts or increments the duplicated block in the time domain, while equation 50 circle shifts data blocks in the frequency domain. The mapper ensures that the same data blocks are orthogonal in the frequency domain and that the minimum spacing between blocks is maximized by interleaving the duplicated blocks using even interleaving distances. The dispersive distribution makes the signal less susceptible to burst distortions. Figure 4. 1(b) and (c)

show CMC operation in the presence of frequency-burst interference and time-burst interference. The black area represents the distortion due to interference. Without CMC, data blocks in the frequency domains subject to frequency interference will suffer severe performance degradation. The mapping ensures that there are surviving, unaffected frequency duplicate data blocks that can be recovered, as shown in the circled area in Figure 4. 1(b). For example, although $s_{3,1}$ is distorted, $s_{3,2}$ can be used to recover s_3 .

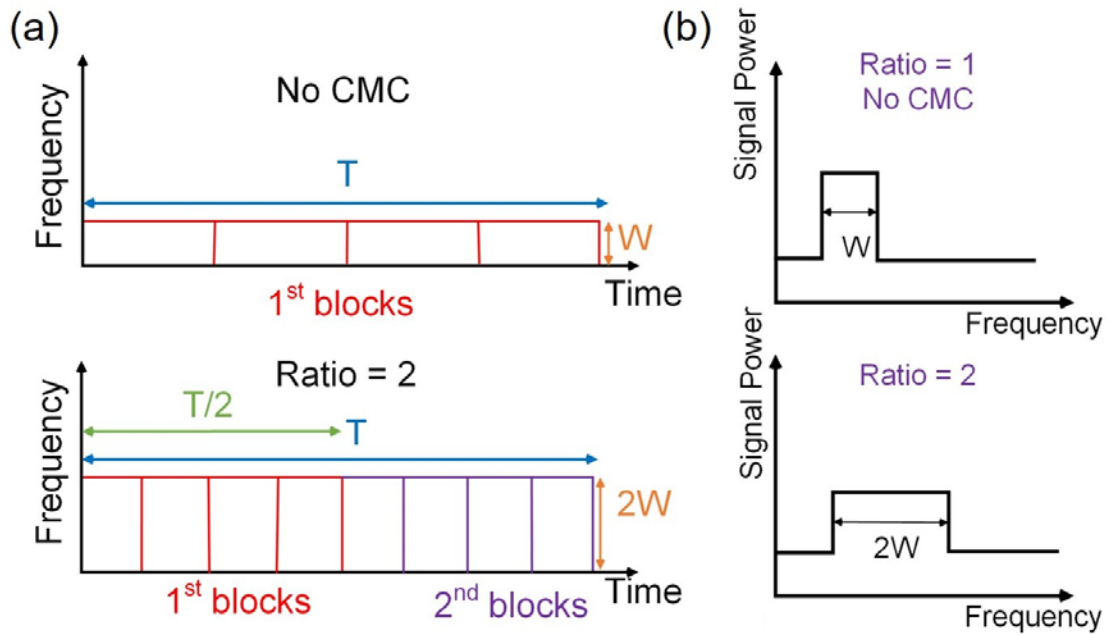


Figure 4. 2: Signal with CMC in the frequency and time domains. ©IEEE 2018

Then we can use the corresponding coordinated combining techniques to recover the distorted blocks with the duplicated information in other frequency locations, thus showing how ample bandwidth can be the trade-off for increased link reliability. Resiliency to time-burst interference can be treated the same way, as shown in Figure 4. 1(c). Let the original time duration per data block be T/N_t , due to a higher sampling rate for a higher mapping ratio M , the time duration of every data block reduces to $T/N_t M$ as depicted in

Figure 4. 2(a). The number of affected or degraded data blocks increases with the constant duration of the time-burst interference. Thus, in Figure 4. 1(c), the number of affected data blocks are doubled in the time domain.

Resiliency benefits are obtained at the expense of occupying more bandwidth and lower signal power assigned to each block for a fixed signaling power. As depicted in Figure 4. 2, as the mapping ratio increases, the required bandwidth increases with increasing sampling rates because the block time duration remains unchanged. Thus, the required bandwidth increases with increasing sampling rate. If the original bandwidth is W , then the bandwidth becomes $M \times W$ for mapping ratio M . The bandwidth cost is not the issue since both mmW link and FSO link have the sufficient bandwidth. Another cost is shown in Figure 4. 2(b). With a constant signal power and white background noise, a higher mapping ratio means more data blocks and proportionally lower signal power and lower signal to noise ratio for each block. Thus, in Figure 4. 2(b), the amplitude of the spectrum decreases with rising ratios. There is a trade-off between mapping ratio and SNR for fixed signaling power. However, under ordinary circumstances there remains the option of increasing the signal power at the expense of overall power consumption.

Coordinated combining is the inverse process of coordinated mapping. It adaptively recovers the repeated blocks in both the mmW and FSO links according to the maximal ratio combining principle [121]. Let the mapping ratios for mmW and FSO be M_1 and M_2 , respectively. Both links are combined in the same data blocks using the following equations:

$$S_{MMW,i}(\Delta t_i, \Delta f_i) = \sum_{k=1}^{M_1} \alpha_{i,k} \left(S_{MMW,i,k}(\Delta t_{i,k}, \Delta f_{i,k}) \right), \quad (51)$$

$$\alpha_{i,k} = \frac{\frac{1}{EVM_{MMW,i,k}^2}}{\sum_{k=1}^{M_1} \frac{1}{EVM_{MMW,i,k}^2}}, \quad (52)$$

$$S_{FSO,i}(\Delta t_i, \Delta f_i) = \sum_{k=1}^{M_2} \beta_{i,k} (S_{FSO,i,k}(\Delta t_{i,k}, \Delta f_{i,k})), \quad (53)$$

$$\beta_{i,k} = \frac{\frac{1}{EVM_{FSO,i,k}^2}}{\sum_{k=1}^{M_1} \frac{1}{EVM_{FSO,i,k}^2}}. \quad (54)$$

EVM is equivalent to $1/\sqrt{SNR}$ and can be calculated from blind EVM estimation [122]. $\Delta t_{i,k}$ and $\Delta f_{i,k}$ can be obtained from equation 49 and 50, $S_{MMW,i,k}$ and $S_{FSO,i,k}$ are the k th mapped block of the i th original block, $\alpha_{i,k}$ and $\beta_{i,k}$ are the weights for $S_{MMW,i,k}$ and $S_{FSO,i,k}$, respectively. $S_{MMW,i}$ and $S_{FSO,i}$ are the i th recovered blocks from the combining operation. $EVM_{MMW,i,k}$ and $EVM_{FSO,i,k}$ are calculated from blind EVM estimation. Since the weights in equation 52 and 54 are inversely proportional to the calculated EVM, they can optimize the system performance adaptively. After coordinated combining, ADCT is used to combine the mmW symbols and FSO symbols to get the final QAM symbols.

4.1.2 Experimental Setup

Figure 4. 3 depicts the experimental setup. Given a bandwidth of 576 MHz and a mapping ratio M , the transmitted 16QAM OFDM signal has a bandwidth of $576 \times M$ MHz with $480 \times M$ -kHz subcarrier spacing, 3.125% CP and 2-Gbps data rate. At the Tx, the 16QAM symbols are split into two branches and mapped independently using coordinated

mapper, followed by OFDM modulation. Then the OFDM symbols are fed into a DAC for the electrical signal generation. For FSO, we employ the IM-DD scheme by using a DML. For mmW, a 1550nm DFB laser is used as light source, which is modulated by a Mach–Zehnder modulator (MZM) driven sinusoidally by a 28.8-GHz microwave source to produce a mmW carrier on a subsequent photodiode. A second MZM is used to modulate data onto the mmW carrier. After transmission via a 21-km SSMF, one branch of the signal is detected by a PD. An antenna with 25-dBi gain paired with a power amplifier (PA) is used for mmW boosting and transmission. The other branch of the signal is coupled to a collimator for FSO transmission. At Rx in mmW, a symmetric antenna, a low-noise amplifier (LNA) and ED are employed to detect and down-convert the mmW signals. A real-time scope samples and stores the signals for offline DSP. In FSO link, a collimator and a PD with 10-GHz bandwidth focus and detect the optical signal. In offline-DSP, the sampled data from the mmW and FSO links is first synchronized to remove the path delays. Then the signals are separately de-mapped using coordinated combining and demodulated to recover the QAM symbols. Finally, we apply the ADCT to combine the mmW and FSO signals together for final EVM measurement and performance analysis.

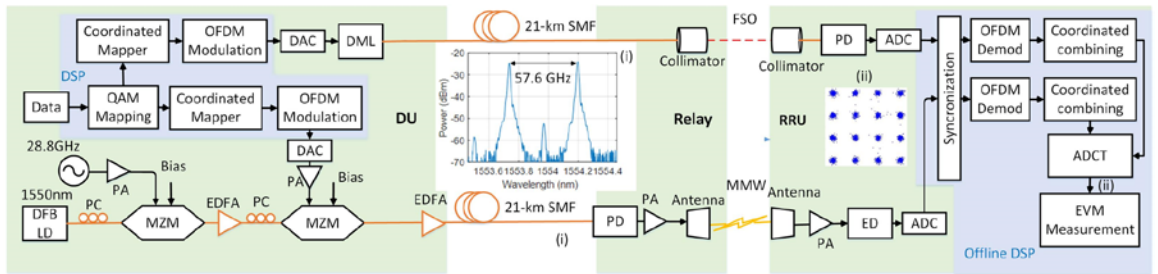


Figure 4. 3: Experimental setup of the mmW/FSO integrated RoF system. ©IEEE 2018

4.1.3 Experimental Results

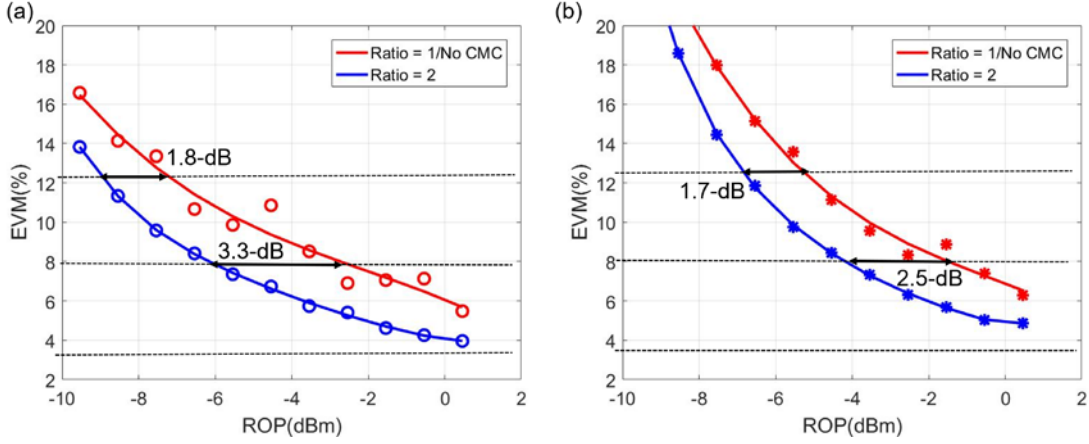


Figure 4. 4: Experimental results of mmW links in (a) BtB experiment and (b) after 21-km SMF. ©IEEE 2018

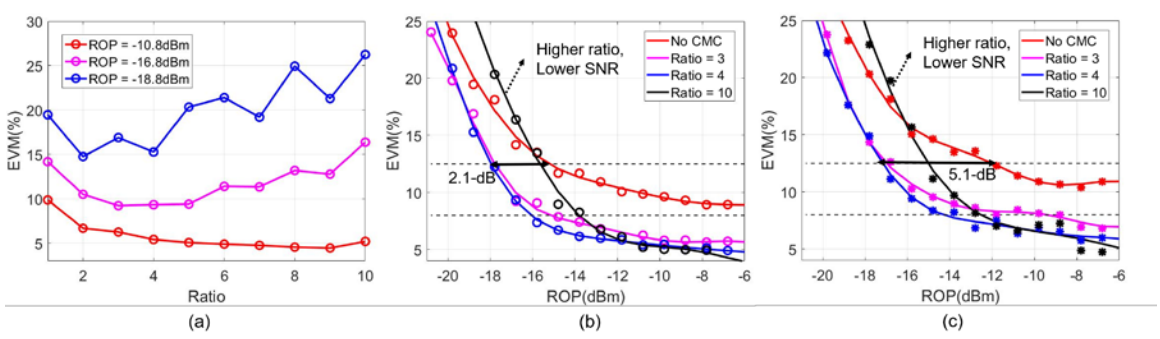


Figure 4. 5: Experimental results of FSO links. (a) EVM performance versus mapping ratio. Power sensitivity performance for three mapping ratios (b) BtB measurements and (c) after 21-km SMF. ©IEEE 2018

As a first step, we investigate the trade-off between mapping ratios and SNR per data block as shown in Figure 4. 5 and Figure 4. 5. Next, we apply the optimal mapping ratios to both links and evaluate the hybrid system's performance in Figure 4. 6. Due to the limited bandwidth of the ED in this experiment, the mmW can only support the mapping ratio up to two. Reference lines of 12.5%, 8% and 3.5% EVM are presented. Figure 4. 4(a) and (b) show the results of power sensitivity measurements in BtB and after 21-km SMF transmission for two different mapping ratios, respectively. The 21-km SMF transmission generates around 2-dB power penalty. Figure 4. 4(b) shows the mmW signal achieves 1.7-

dB/2.5-dB power sensitivity gains with 21-km fiber concerning 12.5%/8% EVM when $M_1 = 2$.

The FSO link has a PD-limited bandwidth of 10 GHz and can support a large number of mapping ratios. Figure 4. 5(a) presents the EVM performance versus mapping ratio from 1 to 10 for three values of ROP, -10.8 dBm, -16.8 dBm, and -18.8 dBm. At higher ROP, the EVM performance is better, particularly for higher mapping ratios. However, with the decreasing ROP, the performance at higher ratios degrades rapidly. The degradation can be attributed to the lower available power, and consequently, the lower SNR ratio per block as the mapping ratio increases as discussed in Figure 4. 2. Therefore, the optimal mapping ratio is 3 and 4. However, in the system with a dynamic gain to maintain a constant SNR over ROP, the optimal mapping ratio can be extended at the expense of power dissipation. Figure 4. 5(b) and (c) show the power sensitivity performances with/without mapping at three ratios. As in Figure 4. 5, higher mapping ratios are more sensitive to ROP because of lower SNR per block. By applying the optimal ratio of CMC, the power sensitivity shows 2.1-dB and 5.1-dB improvement in BTB and 21-km SMF transmission experiments. Moreover, the CMC-based signal maintains an 8% EVM floor down to a ROP about -16 dBm while FSO without CMC is not able to achieve 8% EVM. Therefore, by applying CMC, even an FSO link without mapping can still achieve the 8% EVM requirement.

Using the optimal mapping ratio obtained from a training sequence, we can measure the performance of the hybrid mmW/FSO A-RoF system. The ratios of mmW and FSO are two and four, respectively. Figure 4. 6(a) shows the minimum ROP required to achieve 8%

EVM for both the FSO and mmW links. The single mmW link should satisfy the 8% EVM requirement at -3dBm . The CMC-based hybrid system shows a significant gain compared with the hybrid system without CMC. The ROP tolerance of the mmW link improves 3 dB, 3.6 dB, 4.3 dB and up to 9 dB when the ROP of the FSO link is -21.8 dBm , -19.8 dBm , -17.8 dBm , and -15.8 dBm , respectively. Figure 4. 6(b) shows the power sensitivity performance for a single mmW link and the hybrid mmW/FSO link. Without CMC, the hybrid system can maintain the EVM level below 8% but cannot achieve 3.5% EVM. The CMC-based hybrid system can achieve 3.5% EVM when the ROP of the mmW link is higher than -4 dBm . Its EVM floor can reach 4.2% with an improvement of up to 5.8-dB compared to the hybrid system without CMC.

There are mainly two types of interferences in the system. 1) A mmW link within an unlicensed frequency band, may suffer from out-of-band interference from an adjacent mmW channel leading to frequency burst interference or time burst interference. 2) The FSO channel is robust against in-band stray light interference in part due to the narrowness of its beam and in part owing to narrow band-pass optical filtering but is susceptible to weather conditions such as atmosphere turbulence, scattering and shading or path blockage, which cause power attenuation or fluctuations. To evaluate the robustness of the hybrid system, we assume an undisturbed ROP and add a simulated wireless interference. Here we simulate two cases of mmW link interference. One is a frequency interference from a mmW channel that transmits a 20-MHz long-Term Evolution (LTE) signal and the other is a time burst interference that lasts for around $1/6$ of the single-symbol duration. For FSO, we can adjust the attenuator to achieve power attenuation, which is equivalent to the complicated weather or atmospheric interference.

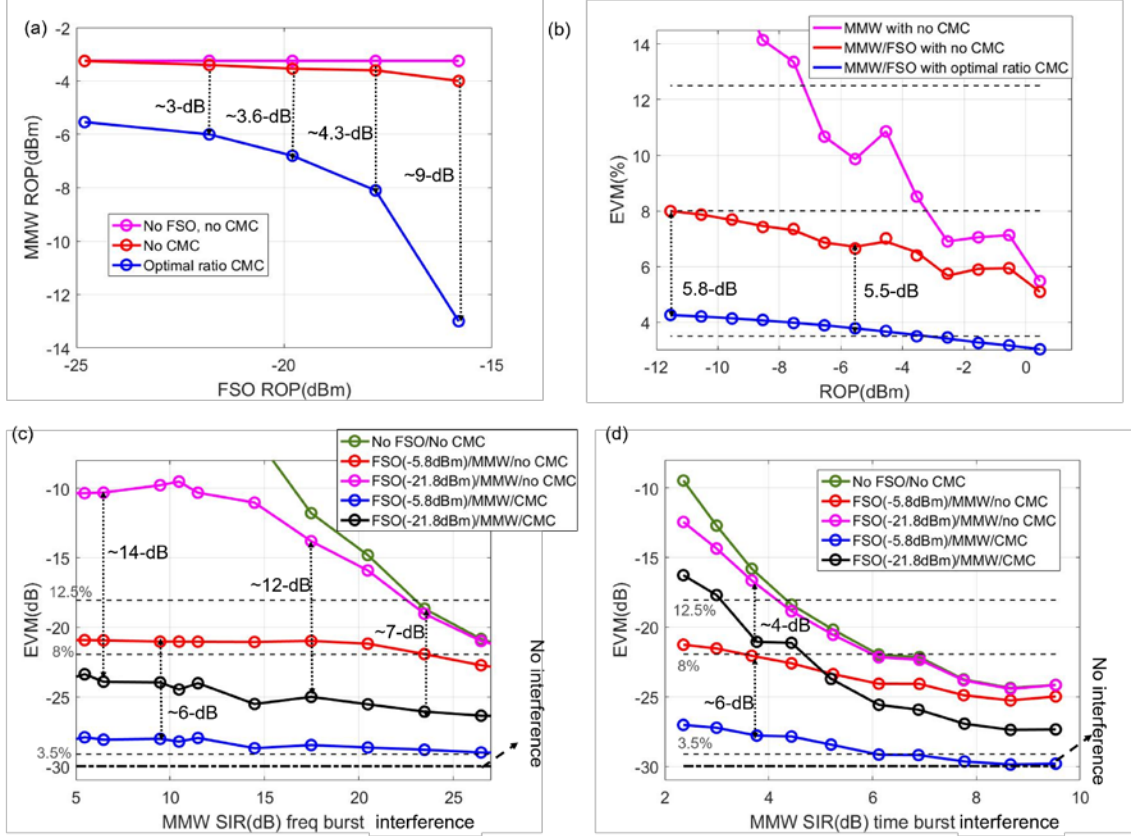


Figure 4. 6: Experimental results of the hybrid mmW and FSO RoF link. ©IEEE 2018

Figure 4. 6(c) shows the EVM performance versus signal-to-interference ratio (SIR) of the system when mmW is subjected to frequency burst interference. If the FSO is in good channel condition, the CMC shows about 6-dB EVM gain compared to the absence of CMC. The CMC-based system shows less than 2-dB EVM penalty from the no interference dashed reference line. Moreover, the EVM floor is maintained at around 4.2%. The worst case occurs when the mmW is under frequency burst attack and the ROP in the FSO channel is minimal. (e.g., ROP = -21.8 dBm). In this case, the performance of the hybrid system without CMC is severely degraded while that of the CMC-based system maintains EVM below the 8 % reference line. The CMC-based system generates around a

2-dB penalty compared to the no interference reference line. It achieves 6-dB to 14-dB EVM gain compared to its counterpart without CMC. As illustrated in Figure 4. 7 (b), the constellation of the recovered QAM signal will be distorted. By utilizing CMC, a good constellation is recovered.

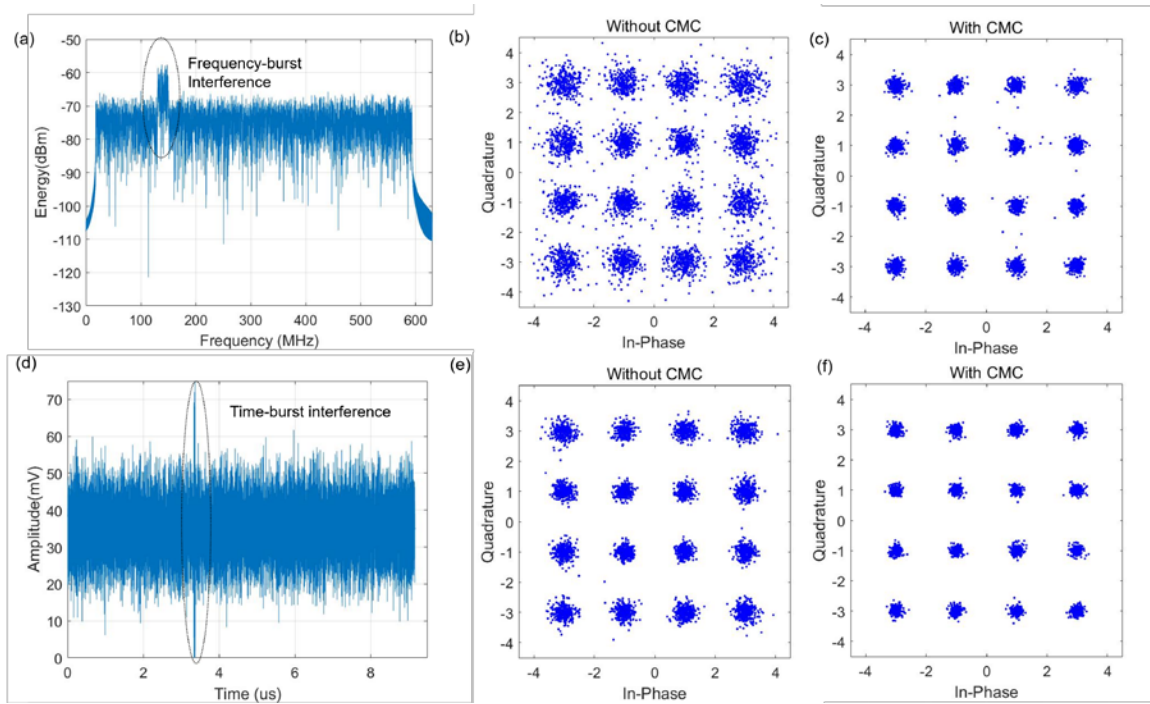


Figure 4. 7: Transmitted OFDM signal and recovered QAM signal with frequency/burst interference. ©IEEE 2018

Figure 4. 6(d) illustrates the EVM performance with SIR variations when the mmW channel is affected by time-burst interference. As shown in Figure 4. 7(d), the interference lasts for 1/6 of a symbol duration. Similar to Figure 4. 6(c), the CMC-based hybrid system shows up to 6-dB EVM gain and generates less than the 3-dB penalty when FSO is in good channel condition. It can tolerate 6-dB SIR while keeping the EVM floor as 3.5%. In the worst case, the CMC-based system achieves around 4-dB EVM gain compared with the non-CMC supported system, which almost follows the curve of the

single mmW link. Figure 4. 7(e) and (f) show the constellations of the affected signals after recovery with and without CMC. Without CMC, time-burst interference leads to a worse constellation. With CMC technique, the constellation is better.

4.1.4 *Summary*

We have proposed a novel CMC technique for fiber-wireless signal transport and experimentally demonstrated improved reliability in a hybrid mmW/FSO A-RoF system. The CMC system exploits the abundant bandwidth of the transmitted OFDM signals and maps each data block to several separated frequency and time locations using a coordinated mapper. A coordinated combining technique combines every duplicated data block with an optimized weight in both FSO and mmW links independently. In the experiment, the CMC-based system shows enhanced power sensitivity and enhanced robustness to frequency/time-burst interference thanks to the dispersive distribution of the repeated data blocks. We show a 9-dB ROP tolerance gain at 8% EVM for the hybrid mmW/FSO system.

Even in extreme operating condition (when the mmW signal is under frequency-burst interference and the FSO channel is severely attenuated), the EVM performance of the CMC-based system only exhibits 1-dB to 2-dB penalty while the system without CMC does not work at all. Thus, the proposed CMC-based A-RoF system achieves a milestone accomplishment for an ultra-reliable and more secure 5G and beyond mobile fronthaul with enhanced system power margins.

CHAPTER 5. CONCLUSION

In this dissertation, we have addressed the flexibility, efficiency, reliability issues in heterogeneous high-capacity communications. The results contribute to both next-generation data center communications and access networks for 5G and beyond applications. We summarize the technical contributions and future research directions as follows.

5.1 Flexibility under Diverse Channel Conditions

5.1.1 Technical Contributions

We have leveraged shaping for different communication systems to enable flexibility under different channel conditions such as SNR, channel loss. Different channel conditions need different probability distributions to optimize signal mutual information. In particular, we have investigated shaping for multi-carrier systems, IM-DD systems with pre amplifiers, and PON.

First, we introduced shaping to enable entropy allocation in OFDM based fiber-mmW systems. In such case, the signal power is determined by the second moment of signal levels, and thus the optimal distribution is bi-polar MB distribution. PAS scheme can be employed to combine with FEC parity bits, where the parity bits are the MSB and determine the signal sign. We have derived a closed-form expression for the SER of probabilistically shaped QAM signal using SNR, under additive AWGN channels. The closed-form expression guides the allocation scheme. The modeling procedure applies to different detection schemes such as ML detector and MAP detector. Then, we reformulate

the PS-unit based entropy allocation scheme to an optimization problem with pre-measured SNR, while the constraint that the averaged entropy and spectral efficiency of the whole data frame remain unchanged. A projected mirror descent algorithm is utilized to maximize the averaged SER among PS units with good convergence rates. We also conduct a proof-of-concept experiment in a fiber-mmW integrated fronthaul link to improve the system performance under uneven frequency response.

As for the single-carrier IM-DD system, the optical signal power is determined by the first moment of signal levels. In this case, the optimal distribution is changed to the unipolar exponential distribution to maximize the entropy. Different from PAS scheme, the parity bits are inserted into the LSB to form a pairwise distribution with negligible performance degradation in certain signal-to-noise ratio ranges. We experimentally demonstrate a LDPC coded PAM-4/8 transmission using the unipolar, pairwise exponential distribution in a fiber-FSO link. Moreover, we employ nonlinear pre-distortion to further mitigate the transmitter nonlinearity. The receiver side utilizes a simple linear equalizer. We investigate two comparison schemes in the experiment. Unipolar distribution exhibits up to 2.1-dB sensitivity improvement in the case of fixed FEC code rate and IR with adjustable signal baud rate. In the case of the same SE by tuning the FEC code rate and signal entropy, the unipolar signal demonstrates up to 1.4-dB ROP gain. Thus, the technique is promising to extend the reach of broadband IM-DD systems.

To allow for more flexibility to support the disaggregation of network functions in the optical transport layer and to better utilize resources, we investigate how to flexibly adapt the data rate of next generation PON systems depending on channel conditions of the users. Since PON has specific requirements including large power budgets (min 29 dB)

and using APDs, the interaction with noise (e.g., shot noise and thermal noise) and link budget optimization need to be studied. In this case, there is no closed-form solution, and we formulate an optimization problem subject to various system constraints and noises. In particular, we maximize the signal mutual information by adapting the shaping settings. We have verified that GS can effectively mitigate the penalty brought by combination with FEC parity bits. In the case of flexible FEC code rate adjusting, the proposed scheme demonstrated up to 1.1-dB sensitivity gain and 18% net data rate improvement compared with uniform PAM4. PS PAM8 gives the optimal GMI among typical PON operating ranges. Monte Carlo simulation using practical LDPC also proved its feasibility to achieve flexible SE tuning with a fixed FEC code rate. The proposed scheme is promising to enable more flexibility for next-generation PON.

5.1.2 Future Directions

While we have proved that shaping enables more flexibility in various system conditions, there are still more research efforts needed to apply the schemes to real-world systems.

First, DM is a critical function block to generate the shaped sequence from binary bits. However, most DMs are power-consuming and need a long encode block length to ensure a small rate loss [123]. Lookup table-based DM is cost efficient but lacks flexibility of arbitrary entropy tuning. We need to design a DM algorithm with higher power efficiency, better rate loss performance, and flexible entropy adjusting. Recently, there are some research efforts using shell mapping [32], multiset-partition DM [30], etc. Real-time

investigation and demonstration of a reconfigurable DM design are critical for practical implementation.

Second, while we verified that GS can compensate for the penalty from the combination with FEC parity bits, the results are based on high-resolution ADC. The impact of low-resolution ADC on the GS signal needs to be investigated. Third, this dissertation focuses on leveraging shaping to specific system models. A proper machine learning model is promising to adapt shaping techniques to different system models and more complicated application scenarios.

5.2 Efficiency Analysis and Efficient System Design

5.2.1 Technical Contributions

To improve system efficiency, we have conducted both trade-off analysis as well as efficient system and algorithm design. In particular, we have performed trade-off analysis between the laser phase noise and signal performance in inter-data center and long-haul transmissions. Then we investigated the cost-effective self-homodyne coherent system design for intra-data center communication. Also, we have proposed a data-efficient nonlinearity correction algorithms by leveraging semi-supervise method and Lasso.

For inter-data center and long-haul communications, tunable lasers are critical to coherent optical transmission. The laser phase noise introduces not only phase variation but also EEPN due to the intermixing with fiber dispersion. EEPN causes significant ISI with long fiber distance, and it is hard to be mitigated through DSP. Thus, we have analyzed the impact of white frequency noise and frequency jitter on different coherent systems. The

theoretical analysis gives a closed-form expression of its impact on signal EVM, which also depends on fiber length and baud rate. We also validated the theoretical result using simulation and experimental demonstration. Moreover, we also characterize the impact of laser flicker noise on the beyond 600G system performance. All these results indicate the importance of setting up laser phase noise masks that are discussed in this dissertation, to qualify tunable lasers and transceivers based on the system operating conditions.

As for intra-data center, since its transmission distance is quite short ($<2\text{km}$) and it is cost-sensitive, we have investigated cost-effective system architecture design with simplified DSP. We have experimentally shown that a 120-Gbaud DP-16QAM self-homodyne coherent system (a net data rate of $>800\text{Gb/s}$) can achieve link budgets of 4.5 dB (for amplified 16 DWDM channels, which can be extended to 6.5dB if 23dBm EDFAs were used) and 5.2 dB (unamplified single-channel case) via a low-latency FEC and a simplified 4×2 MIMO DSP algorithm. The simplified 4×2 MIMO EQ can correct (a) fiber chromatic dispersion for at least 1km, (b) the random I-Q rotation angle caused by the length mismatch between the signal and remote LO paths, and (c) the transceiver I&Q impairments. The amplified DWDM can greatly enhance the link capacity in comparing to the un-amplified case. Thus, C-band DWDM SHD transceivers are very promising for future intra-data center applications.

Moreover, to improve the data efficiency of the conventional nonlinearity correction method implemented in broadband systems, we have introduced semi-supervised method and Lasso to the VNLE. Experimental results in a fiber-FSO link validate that Lasso can reduce the required pilot symbol number by exploiting the sparsity of the tap coefficients. Our experimental results have shown that, in both the 50-Gbaud PAM4 transmission (at

the 1.22×10^{-2} threshold) and 35-Gbaud PAM8 transmission (at the 2×10^{-2} threshold), the semi-supervised sVNLE and Lasso demonstrates ROP gains up to 4-dB and 3-dB over the supervised sVNLE using OLS and the supervised sVNLE using Lasso, respectively. Furthermore, the proposed methods show significant line rate improvements and training overhead reductions over existing techniques at certain BER thresholds.

5.2.2 Future Directions

In this dissertation, we have experimentally demonstrated the 800Gbits/s self-homodyne coherent optical system with simplified offline DSP designs for next-generation intra-data center communications. For instance, we have de-rotated the constant phase rotation caused by the path length difference between the signal and LO in the digital domain (after ADC). Through preliminary study, we have found that the digital phase rotation method combined with the receiver I&Q skew introduces additional penalty to the system. Designing an effective phase de-rotation method in analog domain before ADC is promising to further improve the system performance. Also, we have preliminary verified the feasibility of reusing the dither tone in bias control at the modulator to track the signal polarization through offline processing. Implementing and validating the polarization tracking in a real-time, and analog fashion is critical for the future implementation. How to design the coordination between polarization tracking scheme and automatic bias control (e.g., interleaving the two methods in time) is also important.

Finally, in this dissertation, we only investigated PON using IM-DD scheme. Coherent PON has attracted many research attentions during recent years. PON has longer transmission distance (tens of kilometres) and more stringent power budget than intra-data

center communications. Similar to the design of intra-data center communications, how to simplify the coherent systems in PON is one critical topic in coherent PON.

5.3 Reliability in Fiber-Wireless Integrated Links

5.3.1 Technical Contributions

We propose and experimentally demonstrate a system which integrates mmW and FSO analog radio-over-fiber channels and provides enhanced reliability in 5G and beyond mobile fronthaul links. Given that mmW and FSO are sensitive to different system impairments, the integrated system was implemented by adaptively combining the two complementary links, mmW and FSO, to enhance the reliability of the hybrid wireless system under different channel conditions. The integrated systems utilize a novel CMC technique that senses the signal quality in each data block and apportions the load accordingly. Furthermore, the technique takes advantage of the abundant bandwidth in both mmW and FSO links and is compatible with OFDM. In the CMC design, the data blocks at the transmitter side are interleaved and repeated in both the frequency domain and the time domain, and then the duplicated blocks are combined with adaptive weight coefficients at the receiver side. In experiments, compared with the ADCT-based system, we find a 9-dB gain in ROP tolerance and a 5.8-dB lower EVM floor under a deliberately injected 20-MHz LTE frequency interference burst.

5.3.2 Future Directions

To make the multiple spectra operation feasible in practice, it is important to integrate different optical and electrical device together through electronics and photonics

integration circuits. Also, it is critical to improve the reliability of single mmW or single FSO link because the multiple-spectra system enhances the robustness at the cost of high power consumption and system cost. For instance, interference cancellation and spectra coordination techniques can be applied to mmW systems. DSP techniques such as space-time coding and FEC are also needed to be revised in FSO link [17], considering different system constraints and atmosphere turbulences.

APPENDIX A. DERIVATION OF LASER PHASE NOISE IMPACT

This appendix illustrates the derivation of I_1 and I_2 in equation 33, chapter 2.

A.1 Derivation of I_1

The following analysis calculates I_1 . For $t = nT_s$, we have $r(t) = c_n$. Since the process is stationary, namely, the distributions of signal and the phase noise do not change with n , without loss of generality, we can assume $n = 0$. Plugging in the expression for $r(t)$ and expanding the square of summation, we obtain:

$$\begin{aligned} \mathbb{E} \frac{|r(t - \frac{kf_1}{\pi})|^2 |r(t - \frac{kf_3}{\pi})|^2}{|r(t)|^2} &= \mathbb{E} \left(\frac{1}{|c_0|^2} \mathbb{E} \left(\left(\sum_m |c_m|^2 s_m^2(f_1) \right) \left(\sum_m |c_m|^2 s_m^2(f_3) \right) |c_0 \right) \right) \\ &\quad + \mathbb{E} \left(\frac{1}{|c_0|^2} \mathbb{E} \left(\sum_{m_1 \neq m_2} |c_{m_1}|^2 |c_{m_2}|^2 s_{m_1}(f_1) s_{m_2}(f_1) s_{m_1}(f_3) s_{m_2}(f_3) |c_0 \right) \right) \quad (55) \\ &= Z_1(f_1, f_3) + Z_2(f_1, f_3) \end{aligned}$$

where we denote $s_n(f) = \text{sinc}(kf / T_s + n\pi)$. By plugging equation 55 into the first term in

equation 33, we can decompose $\mathbb{E} \frac{|r(t - \frac{kf_1}{\pi})|^2 |r(t - \frac{kf_3}{\pi})|^2}{|r(t)|^2}$ into two terms, and write I_1 as:

$$I_1 = \frac{1}{\mathbb{E} |r(t)|^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} S(f_1) S(f_3) (Z_1(f_1, f_3) + Z_2(f_1, f_3) - 1) df_1 df_3 \quad (56)$$

For Z_1 , by changing the order of integration, we have:

$$Z_1(f_1, f_3) = \mathbb{E} \left(\frac{1}{|c_0|^2} \mathbb{E} \left(\left(\sum_m |c_m|^2 s_m^2(f_1) \right) \left(\sum_m |c_m|^2 s_m^2(f_3) \right) |c_0 \right) \right)$$

$$\begin{aligned}
&= \mathbb{E} |c_0|^2 s_0^2(f_1) s_0^2(f_3) + \mathbb{E} |c_m|^2 \left(\sum_{m \neq 0} s_0^2(f_1) s_m^2(f_3) + s_0^2(f_3) s_m^2(f_1) \right) \\
&+ \mathbb{E} \frac{1}{|c_0|^2} \left(\sum_{m \neq 0} |c_m|^2 s_m^2(f_1) \right) \left(\sum_{m \neq 0} |c_m|^2 s_m^2(f_3) \right).
\end{aligned} \tag{57}$$

For Z_2 , direct calculation yields:

$$\begin{aligned}
Z_2(f_1, f_3) &= \mathbb{E} \left[\frac{1}{|c_0|^2} \mathbb{E} \left(\sum_{m_1 \neq m_2} |c_{m_1}|^2 |c_{m_2}|^2 s_{m_1}(f_1) s_{m_2}(f_1) s_{m_1}(f_3) s_{m_2}(f_3) |c_0 \right) \right] \\
&= \mathbb{E} \frac{1}{|c_0|^2} \left(\mathbb{E} |c_0|^2 \right)^2 \left[\left(\sum_m s_m(f_1) s_m(f_3) \right)^2 - \sum_m s_m^2(f_1) s_m^2(f_3) \right] \\
&= \mathbb{E} (|c_0|^2) \left[\left(\frac{\sin(k(f_1 - f_3)/T_s)}{(k(f_1 - f_3)/T_s)} \right)^2 - \sum_m s_m^2(f_1) s_m^2(f_3) \right] \approx \mathbb{E} (|c_0|^2) \left[1 - \frac{1}{3} \left(\frac{k(f_1 - f_3)}{T_s} \right)^2 - \sum_m s_m^2(f_1) s_m^2(f_3) \right]
\end{aligned} \tag{58}$$

The expression of I_1 relies on taking integral of Z_1 and Z_2 under the power spectral density $S(f)$, which involves integrals of squared sinc function. In the following we first analyze the behavior of sinc functions in this problem. For $x \ll 1$, we use the following Taylor expansion of sinc function to approximately evaluate I_1 :

$$\text{sinc}(x) = 1 - \frac{1}{6} x^2 + O(x^4), \quad \text{sinc}(x - m\pi) = \frac{(-1)^m x}{m\pi} + O(x^3). \tag{59}$$

By making the decomposition $kf_i/T_s = q_i + x_i$, with integer q_i and $|x_i| < 1/2$, for $i = 1, 2, 3, 4$. By the Taylor expansion approximation, we have:

$$s_m(f_i) = \begin{cases} \frac{(-1)^{q_i - m}}{(q_i - m)\pi} x_i + O(x_i^3), & q_i \neq m \\ 1 - \frac{1}{6} x_i^2 + O(x_i^4), & q_i = m \end{cases} \tag{60}$$

When applying the power spectral densities as kernels, decomposing kf_i/T_s into $q_i + x_i$ and plugging in the Taylor approximation, we obtain:

$$\begin{aligned} \int_{-\infty}^{\infty} s_m^2(f) S(f) df \approx & \left(\frac{k}{T_s} \right)^2 \sum_{q \neq m} \int_{\frac{T_s}{2k}}^{\frac{T_s}{2k}} \frac{f^2}{\pi^2 (q-m)^2} S\left(f + \frac{T_s}{k} q\right) df \\ & + \int_{\frac{T_s}{2k}}^{\frac{T_s}{2k}} \left(1 - \frac{k^2}{3T_s^2} f^2 \right) S\left(\frac{T_s}{k} m + f\right) df \end{aligned} \quad (61)$$

In the second term, we are using the approximation $(1-x)^2 \approx 1-2x$ for small x . In our problem, we have $T_s/k \ll \Delta f_{pp}, \Delta \nu$ ($T_s/k \in [10^9, 10^{11}]$, $\Delta f_{pp} \in [10^5, 10^8]$, $\Delta \nu \sim 10^5$). As derived in Section 2.1.2, the power spectral density is approximated by a mixture of Lorentzian distributions $S(f) \approx \sum_{n=-\infty}^{\infty} w_n \cdot S_l(f - nf_t)$, with each mixture component centered at scale Δf_{pp} and with width $\Delta \nu$. For $|f| \geq T_s/k$, we can use $S(f) \approx \Delta \nu / \pi f^2$, and the integral within $[-T_s/k, T_s/k]$ will be calculated later using the mixture of Lorentz distributions derived in Section II.A. If $m = 0$, we have:

$$\begin{aligned} \int_{-\infty}^{\infty} s_0^2(f) S(f) df & \approx \frac{T_s}{k} \sum_{q \neq 0} \frac{1}{\pi^2 q^2} S\left(\frac{T_s q}{k}\right) + \int_{\frac{T_s}{2k}}^{\frac{T_s}{2k}} \left(1 - \frac{k^2}{3T_s^2} f^2 \right) S(f) df \\ & \approx 1 - \frac{1}{3} \int_{\frac{T_s}{2k}}^{\frac{T_s}{2k}} (kf/T_s)^2 S(f) df + a_1 \frac{k \cdot \Delta \nu}{T_s} \end{aligned} \quad (62)$$

The value of universal constant a_1 is independent of all other parameters, which can be explicitly calculated, but the specific value is irrelevant to our discussion and is dependent on the DSP algorithm that employed at the receiver side. For the case of $m \neq 0$, we have:

$$\begin{aligned}
& \int_{-\infty}^{\infty} s_m^2(f) S(f) df \\
& \approx \left(\frac{k}{T_s} \right)^2 \sum_{\substack{q \neq m \\ q \neq 0}} \int_{-\frac{T_s}{2k}}^{\frac{T_s}{2k}} \frac{f^2}{\pi^2 (q-m)^2} S\left(\frac{T_s q}{k}\right) df + S\left(\frac{T_s m}{k}\right) \int_{-\frac{T_s}{2k}}^{\frac{T_s}{2k}} \left(1 - \frac{k^2}{3T_s^2} f^2\right) df \\
& \quad + \frac{1}{\pi^2 m^2} \int_{-\frac{T_s}{2k}}^{\frac{T_s}{2k}} (kf / T_s)^2 S(f) df \\
& \approx \frac{1}{\pi^2 m^2} \int_{-\frac{T_s}{2k}}^{\frac{T_s}{2k}} (kf / T_s)^2 S(f) df + a_2 \frac{k \cdot \Delta \nu}{m^2 T_s}.
\end{aligned} \tag{63}$$

Similar to a_1 , a_2 is a universal constant that does not change with other parameters. Going back to I_1 , we need the integral of $Z_1(f_1, f_3)$ under the kernel $S(f_1)S(f_3)$. By plugging in the approximate formulae for each term in the expression of $Z_1(f_1, f_3)$, we obtain:

$$\begin{aligned}
& \frac{1}{\mathbb{E} |c_0|^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} Z_1(f_1, f_3) S(f_1) S(f_3) df_1 df_3 \\
& \stackrel{(i)}{=} \left(\int_{-\infty}^{\infty} s_0^2(f_1) S(f_1) df_1 \right) \left(\int_{-\infty}^{\infty} s_0^2(f_3) S(f_3) df_3 \right) + 2 \sum_{m \neq 0} \left(\int_{-\infty}^{\infty} s_m^2(f_3) S(f_3) df_3 \right) \left(\int_{-\infty}^{\infty} s_0^2(f_1) S(f_1) df_1 \right) \\
& \quad + \left(\sum_{m \neq 0} \int_{-\infty}^{\infty} s_m^2(f_1) S(f_1) df_1 \right) \left(\sum_{m \neq 0} \int_{-\infty}^{\infty} s_m^2(f_3) S(f_3) df_3 \right) \\
& \stackrel{(ii)}{\approx} \left(1 - \frac{1}{3} \int_{-\frac{T_s}{2k}}^{\frac{T_s}{2k}} (kf_1 / T_s)^2 S(f_1) df_1 + a_1 \frac{k \cdot \Delta \nu}{T_s} \right)^2 \\
& \quad + 2 \left(\sum_{m \neq 0} \frac{1}{\pi^2 m^2} \int_{-\frac{T_s}{2k}}^{\frac{T_s}{2k}} (kf / T_s)^2 S(f) df + a_2 \frac{k \cdot \Delta \nu}{m^2 T_s} \right) \left(1 - \frac{1}{3} \int_{-\frac{T_s}{2k}}^{\frac{T_s}{2k}} (kf_1 / T_s)^2 S(f_1) df_1 + a_1 \frac{k \cdot \Delta \nu}{T_s} \right) \\
& \quad + \left(\sum_{m \neq 0} \frac{1}{\pi^2 m^2} \int_{-\frac{T_s}{2k}}^{\frac{T_s}{2k}} (kf / T_s)^2 S(f) df + a_2 \frac{k \cdot \Delta \nu}{m^2 T_s} \right)^2 \\
& \stackrel{(iii)}{\approx} 1 - \frac{2}{3} \int_{-\frac{T_s}{2k}}^{\frac{T_s}{2k}} (kf_1 / T_s)^2 S(f_1) df_1 + a_1 \frac{k \cdot \Delta \nu}{T_s} + \left(\sum_{m \neq 0} \frac{2}{\pi^2 m^2} \right) \int_{-\frac{T_s}{2k}}^{\frac{T_s}{2k}} (kf / T_s)^2 S(f) df + a_2 \sum_{m \neq 0} \frac{k \cdot \Delta \nu}{m^2 T_s} \\
& \stackrel{(iv)}{=} 1 + \left(a_1 + \frac{\pi^2}{3} a_2 \right) \frac{k \cdot \Delta \nu}{T_s}.
\end{aligned} \tag{64}$$

The above calculation involves four steps. Step (i) is by exchanging the order of summation and differentiation; step (ii) is by plugging in the approximate formulae for the squared

integral of sinc functions; step (iii) is discarding high-order terms when the integrals are

less than 1, and step (iv) is using the identity $\sum_{n \geq 1} n^{-2} = \frac{\pi^2}{6}$.

In the above calculation, we omit all the high order terms. The last squared term in the expression is negligible because it is of the same order of square of other terms.

For $Z_2(f_1, f_3)$, we have:

$$\begin{aligned} & \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} Z_2(f_1, f_3) S(f_1) S(f_3) df_1 df_3 \\ & \approx (\mathbb{E} |c_0|^2) \cdot \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left[1 - \frac{1}{6} \left(\frac{k(f_1 - f_3)}{T_s} \right)^2 - \sum_m s_m^2(f_1) s_m^2(f_3) \right] S(f_1) S(f_3) df_1 df_3 \quad (65) \\ & \approx (\mathbb{E} |c_0|^2) \cdot \left(1 - \sum_m \left(\int_{-\infty}^{\infty} s_m^2(f) S(f) df \right)^2 \right) \approx a^2 (\mathbb{E} |c_0|^2) \left(\frac{k \cdot \Delta \nu}{T_s} \right)^2 \end{aligned}$$

Since $k \cdot \Delta \nu / T_s \ll 1$, the integral of Z_2 is much smaller than $k \cdot \Delta \nu / T_s$, which is the scale of the integral of Z_1 . Therefore, the contribution by Z_2 in the final expression of I_1 is negligible, and we obtain:

$$I_1 \approx a' \frac{k \cdot \Delta \nu}{T_s}, \quad (66)$$

for some constant $a' > 0$.

A.2 Derivation of $I_2 - I_1$

Now we turn to the term I_2 . Note that:

$$I_2 - I_1 = \frac{1}{4\mathbb{E}|c_0|^2} \iint_{\square^2} S(f_1)S(f_3) \mathbb{E} \left| r(t - \frac{kf_1}{\pi}) - r(t - \frac{kf_3}{\pi}) \right|^2 df_1 df_3 \quad (67)$$

Note that for $f_1, f_3 \in T_s/k$, we have:

$$\begin{aligned} \mathbb{E} \left| r(t - \frac{kf_1}{\pi}) - r(t - \frac{kf_3}{\pi}) \right|^2 &= \mathbb{E}(|c_0|^2) \sum_m (s_m(f_1) - s_m(f_3))^2 \\ &= 2\mathbb{E}(|c_0|^2) \left(1 - \frac{\sin(k(f_1 - f_3)/T_s)}{k(f_1 - f_3)/T_s} \right) \approx \frac{2}{3} \mathbb{E}(|c_0|^2) \left(\frac{k(f_1 - f_3)}{T_s} \right)^2. \end{aligned} \quad (68)$$

And as in the estimate for I_1 , the integral outside the region $[-T_s/2k, T_s/2k] \times [-T_s/2k, T_s/2k]$ is approximately a constant multiple of $k \cdot \Delta\nu / T_s$.

Therefore, we have:

$$\begin{aligned} I_2 - I_1 &\approx a \frac{k \cdot \Delta\nu}{T_s} + \int_{-\frac{T_s}{2k}}^{\frac{T_s}{2k}} \int_{-\frac{T_s}{2k}}^{\frac{T_s}{2k}} (k(f_1 - f_3)/T_s)^2 S(f_1)S(f_3) df_1 df_3 \\ &= a \frac{k \cdot \Delta\nu}{T_s} + \int_{-\frac{T_s}{2k}}^{\frac{T_s}{2k}} \int_{-\frac{T_s}{2k}}^{\frac{T_s}{2k}} \frac{k^2}{T_s^2} (f_1^2 + f_3^2 - 2f_1 f_3) S(f_1)S(f_3) df_1 df_3 \\ &\approx a \frac{k \cdot \Delta\nu}{T_s} + 2 \int_{-\frac{T_s}{2k}}^{\frac{T_s}{2k}} \frac{k^2 f^2}{T_s^2} S(f) df \end{aligned} \quad (69)$$

The cross term integrates to zero because the power spectral density is a symmetric function.

REFERENCES

- [1] "Ericsson Mobility Report " November 2021.
- [2] X. Zhou, R. Urata, and H. Liu, "Beyond 1 Tb/s Intra-Data Center Interconnect Technology: IM-DD OR Coherent?," *Journal of Lightwave Technology*, vol. 38, pp. 475-484, 2020.
- [3] Y.-W. Chen, R. Zhang, C.-W. Hsu, and G.-K. Chang, "Key Enabling Technologies for the Post-5G Era: Fully Adaptive, All-Spectra Coordinated Radio Access Network with Function Decoupling," *IEEE Communications Magazine*, vol. 58, pp. 60-66, 2020.
- [4] V. Houtsma, D. v. Veen, and E. Harstead, "Recent Progress on Standardization of Next-Generation 25, 50, and 100G EPON," *Journal of Lightwave Technology*, vol. 35, pp. 1228-1234, 2017.
- [5] G. Chang and L. Cheng, "Fiber-wireless integration for future mobile communications," in *2017 IEEE Radio and Wireless Symposium (RWS)*, 2017, pp. 16-18.
- [6] M. Morsy-Osman, M. Sowailem, E. El-Fiky, T. Goodwill, T. Hoang, S. Lessard, et al., "DSP-free 'coherent-lite' transceiver for next generation single wavelength optical intra-datacenter interconnects," *Optics Express*, vol. 26, pp. 8890-8903, 2018/04/02 2018.
- [7] M. S. Faruk and K. Kikuchi, "Frequency-domain adaptive equalizer with rational oversampling rates in coherent optical receivers," in *2014 The European Conference on Optical Communication (ECOC)*, 2014, pp. 1-3.
- [8] R. Rios-Müller, J. Renaudier, and G. Charlet, "Blind Receiver Skew Compensation and Estimation for Long-Haul Non-Dispersion Managed Systems Using Adaptive Equalizer," *Journal of Lightwave Technology*, vol. 33, pp. 1315-1318, 2015.
- [9] T. Pfau, S. Hoffmann, and R. Noe, "Hardware-Efficient Coherent Digital Receiver Concept With Feedforward Carrier Recovery for M-QAM Constellations," *Journal of Lightwave Technology*, vol. 27, pp. 989-999, 2009.
- [10] G. Liu, K. Zhang, R. Zhang, R. Proietti, H. Lu, and S. J. Ben Yoo, "Demonstration of a carrier frequency offset estimator for 16-/32-QAM coherent receivers: a hardware perspective," *Optics Express*, vol. 26, pp. 4853-4862, 2018/02/19 2018.
- [11] M. S. Faruk and S. J. Savory, "Digital signal processing for coherent transceivers employing multilevel formats," *Journal of Lightwave Technology*, vol. 35, pp. 1125-1141, 2017.
- [12] Z. He, T. Bo, and H. Kim, "Probabilistically shaped coded modulation for IM/DD system," *Optics Express*, vol. 27, pp. 12126-12136, 2019/04/29 2019.

- [13] X. Chen, C. Antonelli, S. Chandrasekhar, G. Raybon, A. Mecozzi, M. Shtaif, et al., "Kramers–Kronig Receivers for 100-km Datacenter Interconnects," *Journal of Lightwave Technology*, vol. 36, pp. 79-89, 2018.
- [14] J. K. Perin, A. Shastri, and J. M. Kahn, "Design of Low-Power DSP-Free Coherent Receivers for Data Center Links," *Journal of Lightwave Technology*, vol. 35, pp. 4650-4662, 2017.
- [15] T. Hirokawa, S. Pinna, N. Hosseinzadeh, A. Maharry, H. Andrade, J. Liu, et al., "Analog Coherent Detection for Energy Efficient Intra-Data Center Links at 200 Gbps Per Wavelength," *Journal of Lightwave Technology*, vol. 39, pp. 520-531, 2021.
- [16] F. Tonini, C. Raffaelli, L. Wosinska, and P. Monti, "Cost-Optimal Deployment of a C-RAN With Hybrid Fiber/FSO Fronthaul," *Journal of Optical Communications and Networking*, vol. 11, pp. 397-408, 2019/07/01 2019.
- [17] A. Trichili, M. A. Cox, B. S. Ooi, and M.-S. Alouini, "Roadmap to free space optics," *Journal of the Optical Society of America B*, vol. 37, pp. A184-A201, 2020/11/01 2020.
- [18] S. Rangan, T. S. Rappaport, and E. Erkip, "Millimeter-Wave Cellular Wireless Networks: Potentials and Challenges," *Proceedings of the IEEE*, vol. 102, pp. 366-385, 2014.
- [19] H. Dahrouj, A. Douik, F. Rayal, T. Y. Al-Naffouri, and M. Alouini, "Cost-effective hybrid RF/FSO backhaul solution for next generation wireless systems," *IEEE Wireless Communications*, vol. 22, pp. 98-104, 2015.
- [20] J. Zhang, J. Wang, Y. Xu, M. Xu, F. Lu, L. Cheng, et al., "Fiber–wireless integrated mobile backhaul network based on a hybrid millimeter-wave and free-space-optics architecture with an adaptive diversity combining technique," *Optics Letters*, vol. 41, pp. 1909-1912, 2016/05/01 2016.
- [21] Y. Jianjun, J. Zhensheng, L. Yi, Y. Su, C. Gee-Kung, and W. Ting, "Optical millimeter-wave generation or up-conversion using external modulators," *IEEE Photonics Technology Letters*, vol. 18, pp. 265-267, 2006.
- [22] J. Yu, Z. Jia, T. Wang, and G. K. Chang, "Centralized Lightwave Radio-Over-Fiber System With Photonic Frequency Quadrupling for High-Frequency Millimeter-Wave Generation," *IEEE Photonics Technology Letters*, vol. 19, pp. 1499-1501, 2007.
- [23] V. E. Houtsma and D. T. v. Veen, "Investigation of Modulation Schemes for Flexible Line-Rate High-Speed TDM-PON," *Journal of Lightwave Technology*, vol. 38, pp. 3261-3267, 2020.
- [24] J. S. Wey, Y. Luo, and T. Pfeiffer, "5G wireless transport in a PON context: An overview," *IEEE Communications Standards Magazine*, vol. 4, pp. 50-56, 2020.

- [25] S. Shen, T. Zhang, S. Mao, and G. K. Chang, "DRL-Based Channel and Latency Aware Radio Resource Allocation for 5G Service-Oriented RoF-MmWave RAN," *Journal of Lightwave Technology*, vol. 39, pp. 5706-5714, 2021.
- [26] G. Böcherer, P. Schulte, and F. Steiner, "Probabilistic Shaping and Forward Error Correction for Fiber-Optic Communication Systems," *Journal of Lightwave Technology*, vol. 37, pp. 230-244, 2019.
- [27] Q. Zhou, S. Shen, Y. W. Chen, R. Zhang, J. Finkelstein, and G. K. Chang, "Simultaneous Nonlinear Self-Interference Cancellation and Signal of Interest Recovery Using Dual Input Deep Neural Network in New Radio Access Networks," *Journal of Lightwave Technology*, vol. 39, pp. 2046-2051, 2021.
- [28] F. R. Kschischang and S. Pasupathy, "Optimal nonuniform signaling for Gaussian channels," *IEEE Transactions on Information Theory*, vol. 39, pp. 913-929, 1993.
- [29] P. Schulte and G. Böcherer, "Constant Composition Distribution Matching," *IEEE Transactions on Information Theory*, vol. 62, pp. 430-434, 2016.
- [30] T. Fehenberger, D. S. Millar, T. Koike-Akino, K. Kojima, and K. Parsons, "Multiset-partition distribution matching," *IEEE Transactions on Communications*, vol. 67, pp. 1885-1893, 2018.
- [31] T. Fehenberger, D. S. Millar, T. Koike-Akino, K. Kojima, K. Parsons, and H. Griesser, "Huffman-coded sphere shaping and distribution matching algorithms via lookup tables," *Journal of Lightwave Technology*, vol. 38, pp. 2826-2834, 2020.
- [32] P. Schulte and F. Steiner, "Divergence-optimal fixed-to-fixed length distribution matching with shell mapping," *IEEE Wireless Communications Letters*, vol. 8, pp. 620-623, 2019.
- [33] Z. Qu and I. B. Djordjevic, "On the probabilistic shaping and geometric shaping in optical communication systems," *IEEE Access*, vol. 7, pp. 21454-21464, 2019.
- [34] X. Li, J. Yu, L. Zhao, K. Wang, C. Wang, M. Zhao, et al., "1-Tb/s Millimeter-Wave Signal Wireless Delivery at D-Band," *Journal of Lightwave Technology*, vol. 37, pp. 196-204, 2019.
- [35] T. Sasai, A. Matsushita, M. Nakamura, S. Okamoto, F. Hamaoka, and Y. Kisaka, "Experimental Analysis of Laser Phase Noise Tolerance of Uniform 256QAM and Probabilistically Shaped 1024QAM," in *2019 Optical Fiber Communications Conference and Exhibition (OFC)*, 2019, pp. 1-3.
- [36] T. A. Eriksson, M. Chagnon, F. Buchali, K. Schuh, S. t. Brink, and L. Schmalen, "56 Gbaud Probabilistically Shaped PAM8 for Data Center Interconnects," in *2017 European Conference on Optical Communication (ECOC)*, 2017, pp. 1-3.

- [37] R. F. Nalewajski, "Elements of information theory," in *Perspectives in Electronic Structure Theory*, ed: Springer, 2011, pp. 371-395.
- [38] R. Zhang, Y. Chen, W. Mou, and G. Chang, "Rate Redundancy and Entropy Allocation for PAS-OFDM based Mobile Fronthaul," *Journal of Lightwave Technology*, pp. 1-1, 2020.
- [39] R. Zhang, Y.-W. Chen, W. Mou, S. Yao, and G.-K. Chang, "Entropy Allocation Optimization for PS-OFDM With Constellation Partitioning Based Modeling," *Journal of Lightwave Technology*, vol. 38, pp. 6024-6030, 2020/11/01 2020.
- [40] R. Zhang, Y.-W. Chen, S. Shen, Q. Zhou, S. Yao, S. J. Su, et al., "Joint Optimization of Processing Complexity and Rate Allocation through Entropy Tunability for 64-/256-QAM Based Radio Fronthauling with LDPC and PAS-OFDM," in *Optical Fiber Communication Conference (OFC) 2020*, San Diego, California, 2020, p. M2F.2.
- [41] R. Zhang, C.-W. Hsu, X. Tang, Q. Zhou, and G.-K. Chang, "LDPC Coded PAM-4/8 Transmission in Fiber-FSO Link Using Unipolar Probability Distribution and Pre-distortion," in *Optical Fiber Communication Conference (OFC) 2021*, Washington, DC, 2021, p. Th5E.5.
- [42] R. Zhang, N. Kaneda, Y. Lefevre, A. Mahadevan, D. v. Veen, and V. Houtsma, "Probabilistic and Geometric Shaping for Next-Generation 100G Flexible PON," in *2020 European Conference on Optical Communications (ECOC)*, 2020, pp. 1-4.
- [43] N. Kaneda, R. Zhang, Y. Lefevre, A. Mahadevan, D. van Veen, and V. Houtsma, "First Experimental Demonstration of Flexible Rate PON Beyond 100Gb/s with Probabilistic and Geometric Shaping," in *Optical Fiber Communication Conference (OFC) 2021*, Washington, DC, 2021, p. F2H.2.
- [44] J. Buus and E. J. Murphy, "Tunable lasers in optical networks," *Journal of Lightwave Technology*, vol. 24, p. 5, 2006.
- [45] M. Iglesias Olmedo, X. Pang, R. Schatz, O. Ozolins, H. Louchet, D. Zibar, et al., "Effective linewidth of semiconductor lasers for coherent optical data links," in *Photonics*, 2016, p. 39.
- [46] W. Shieh and K.-P. Ho, "Equalization-enhanced phase noise for coherent-detection systems using electronic digital signal processing," *Optics Express*, vol. 16, pp. 15718-15727, 2008/09/29 2008.
- [47] I. Fatadin and S. J. Savory, "Impact of phase to amplitude noise conversion in coherent optical systems with digital dispersion compensation," *Optics Express*, vol. 18, pp. 16273-16278, 2010/07/19 2010.
- [48] A. Kakkar, J. R. Navarro, R. Schatz, H. Louchet, X. Pang, O. Ozolins, et al., "Comprehensive Study of Equalization-Enhanced Phase Noise in Coherent Optical Systems," *Journal of Lightwave Technology*, vol. 33, pp. 4834-4841, 2015.

- [49] A. Kakkar, J. R. Navarro, R. Schatz, X. Pang, O. Ozolins, A. Udalcovs, et al., "Laser frequency noise in coherent optical systems: spectral regimes and impairments," *Scientific reports*, vol. 7, pp. 1-10, 2017.
- [50] M. Al-Qadi, G. Vedala, and R. Hui, "Phase Noise of Diode Laser Frequency Comb and its Impact in Coherent Communication Systems," in *Conference on Lasers and Electro-Optics*, San Jose, California, 2018, p. JTu2A.35.
- [51] F. Cruz, R. Cavasso-Filho, A. Siqueira, D. Manoel, E. Telles, A. Scalabrin, et al., "Laser noise measurements and observation of amplitude squeezing in an extended cavity diode laser," *Revista de Física Aplicada e Instrumentaçã o*, vol. 14, 1999.
- [52] D. Derickson, "Fiber optic test and measurement," *Fiber optic test and measurement*/edited by Dennis Derickson. Upper Saddle River, 1998.
- [53] K. Kikuchi, "Characterization of semiconductor-laser phase noise and estimation of bit-error rate performance with low-speed offline digital coherent receivers," *Optics Express*, vol. 20, pp. 5291-5302, 2012.
- [54] T. N. Huynh and L. P. Barry, "Overcoming the effect of cycle slips caused by low frequency noise from monolithic tunable lasers," in *OFC 2014*, 2014, pp. 1-3.
- [55] T. N. Huynh, A. T. Nguyen, W.-C. Ng, L. Nguyen, L. A. Rusch, and L. P. Barry, "BER Performance of Coherent Optical Communications Systems Employing Monolithic Tunable Lasers With Excess Phase Noise," *Journal of Lightwave Technology*, vol. 32, pp. 1973-1980, 2014/05/15 2014.
- [56] H. Ishii, F. Kano, Y. Tohmori, Y. Kondo, T. Tamamura, and Y. Yoshikuni, "Narrow spectral linewidth under wavelength tuning in thermally tunable super-structure-grating (SSG) DBR lasers," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 1, pp. 401-407, 1995.
- [57] J. P. Wilde, G. W. Yoffe, and J. M. Kahn, "Frequency Noise Characterization of a Widely Tunable Narrow-Linewidth DFB Laser Array Source," in *Optical Fiber Communication Conference and National Fiber Optic Engineers Conference*, San Diego, California, 2009, p. JWA33.
- [58] J. E. Simsarian, J. Gripp, S. Chandrasekhar, and P. Mitchell, "Fast-Tuning Coherent Burst-Mode Receiver for Metropolitan Networks," *IEEE Photonics Technology Letters*, vol. 26, pp. 813-816, 2014.
- [59] L. D. Turner, K. Weber, C. Hawthorn, and R. E. Scholten, "Frequency noise characterisation of narrow linewidth diode lasers," *Optics communications*, vol. 201, pp. 391-397, 2002.
- [60] R. Zhang, W. Jiang, Jr., K. Kuzmin, R. Juluri, G.-K. Chang, and W. I. Way, "Laser Frequency Jitter Tolerance and Linewidth Requirement for >64Gbaud DP-16QAM

coherent systems," in Optical Fiber Communication Conference (OFC) 2019, San Diego, California, 2019, p. M4I.2.

[61] R. Zhang, K. Kuzmin, W. Jiang, Jr., G. Giaretta, T. Tomimoto, Y. Weng, et al., "Impact of laser flicker noise and linewidth on 64 to 96 Gbaud/DP-nQAM metro coherent optical links," *Optics Letters*, vol. 45, pp. 1220-1223, 2020/03/01 2020.

[62] R. Zhang, W. Jiang, K. Kuzmin, Y. Weng, W. Mou, G. Chang, et al., "The Impact of Local Oscillator Frequency Jitter and Laser Linewidth to Ultra High Baud Rate Coherent Systems," *Journal of Lightwave Technology*, vol. 38, pp. 1138-1147, 2020.

[63] T. Gui, X. Wang, M. Tang, Y. Yu, Y. Lu, and L. Li, "Real-Time Demonstration of 600 Gb/s DP-64QAM Self-Homodyne Coherent Bi-Direction Transmission with Un-Cooled DFB Laser," in Optical Fiber Communication Conference Postdeadline Papers 2020, San Diego, California, 2020, p. Th4C.3.

[64] J. Cheng, C. Xie, M. Tang, and S. Fu, "A Comparative Study of Intradyne and Self-homodyne Systems for Next Generation Intra-datacenter Optical Interconnects," in 2019 24th OptoElectronics and Communications Conference (OECC) and 2019 International Conference on Photonics in Switching and Computing (PSC), 2019, pp. 1-3.

[65] R. Zhang, Y.-W. Chen, K. Kuzmin, and W. Way, "Intra-Data Center 120Gbaud/DP-16QAM Self-Homodyne Coherent Links with Simplified Coherent DSP," *Optical Fiber Communication Conference*, 2022.

[66] S. P. Yadav and S. C. Bera, "Nonlinearity effect of high power amplifiers in communication systems," in 2014 International Conference on Advances in Communication and Computing Technologies (ICACACT 2014), 2014, pp. 1-6.

[67] G. Khanna, B. Spinnler, S. Calabrò, E. De Man, U. Feiste, T. Drenski, et al., "A Memory Polynomial Based Digital Pre-Distorter for High Power Transmitter Components," in Optical Fiber Communication Conference, Los Angeles, California, 2017, p. M2C.4.

[68] F. Pittalà, M. Schaedler, C. Bluemm, G. Goeger, S. Calabrò, M. Kuschnerov, et al., "800ZR+ DWDM Demonstration over 600km G.654D Fiber Enabled by Adaptive Nonlinear TripleX Equalization," in Optical Fiber Communication Conference (OFC) 2020, San Diego, California, 2020, p. M4K.5.

[69] E. Changsoo and E. J. Powers, "A new Volterra predistorter based on the indirect learning architecture," *IEEE Transactions on Signal Processing*, vol. 45, pp. 223-227, 1997.

[70] S. Zhalehpour, J. Lin, and L. Rusch, "SiP IQ modulator Linearization by memory polynomial pre-distortion model," in 2017 IEEE Photonics Conference (IPC), 2017, pp. 317-318.

- [71] V. Kekatos and G. B. Giannakis, "Sparse Volterra and Polynomial Regression Models: Recoverability and Estimation," *IEEE Transactions on Signal Processing*, vol. 59, pp. 5907-5920, 2011.
- [72] W.-J. Huang, W.-F. Chang, C.-C. Wei, J.-J. Liu, Y.-C. Chen, K.-L. Chi, et al., "93% Complexity Reduction of Volterra Nonlinear Equalizer by ℓ_1 -Regularization for 112-Gbps PAM-4 850-nm VCSEL Optical Interconnect," in *Optical Fiber Communication Conference*, San Diego, California, 2018, p. M2D.7.
- [73] S. Lu, C. Wei, C. Chuang, Y. Chen, and J. Chen, "81.7% Complexity Reduction of Volterra Nonlinear Equalizer by Adopting L1 Regularization Penalty in an OFDM Long-Reach PON," in *2017 European Conference on Optical Communication (ECOC)*, 2017, pp. 1-3.
- [74] C. Chuang, W. Chang, C. Wei, C. Ho, C. Huang, J. Shi, et al., "Sparse Volterra Nonlinear Equalizer by Employing Pruning Algorithm for High-Speed PAM-4 850-nm VCSEL Optical Interconnect," in *2019 Optical Fiber Communications Conference and Exhibition (OFC)*, 2019, pp. 1-3.
- [75] Q. Zhou, F. Zhang, and C. Yang, "AdaNN: Adaptive Neural Network-Based Equalizer via Online Semi-Supervised Learning," *Journal of Lightwave Technology*, vol. 38, pp. 4315-4324, 2020.
- [76] J. Zhang, Z. Jia, M. Xu, H. Zhang, L. A. Campos, and C. Knittle, "High-Performance Preamble Design and Upstream Burst-Mode Detection in 100 -Gb/s/ λ TDM Coherent-PON," in *Optical Fiber Communication Conference (OFC) 2020*, San Diego, California, 2020, p. W1E.1.
- [77] M. T. Dabiri, S. M. S. Sadough, and M. A. Khalighi, "FSO Communication for High Speed Trains: Blind Data Detection and Channel Estimation," in *2018 11th International Symposium on Communication Systems, Networks & Digital Signal Processing (CSNDSP)*, 2018, pp. 1-4.
- [78] R. Zhang, X. Tang, C. W. Hsu, Y. W. Chen, Q. Zhou, S. Shen, et al., "115-Gbps/ λ Fiber/FSO Transmission with Supervised and Semi-supervised Nonlinearity Correction Using Lasso," in *2021 Optical Fiber Communications Conference and Exhibition (OFC)*, 2021, pp. 1-3.
- [79] R. Zhang, X. Tang, C. W. Hsu, Y. W. Chen, and G. K. Chang, "Semi-Supervised and Supervised Nonlinear Equalizers in Fiber-FSO Converged System," *Journal of Lightwave Technology*, vol. 39, pp. 6175-6181, 2021.
- [80] D.-H. Lee, "Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks," in *Workshop on challenges in representation learning*, ICML, 2013.
- [81] X. Zhu and A. B. Goldberg, "Introduction to semi-supervised learning," *Synthesis lectures on artificial intelligence and machine learning*, vol. 3, pp. 1-130, 2009.

- [82] M. E. Khaled, P. Fortier, M. L. Ammari, and S. A. M. Tariq, "Optimal power allocation for underground selective channel at 60 GHz," in 2014 IEEE International Conference on Communications Workshops (ICC), 2014, pp. 736-741.
- [83] B. He and R. Schober, "Bit-interleaved coded modulation for hybrid RF/FSO systems," IEEE Transactions on Communications, vol. 57, pp. 3753-3763, 2009.
- [84] R. Zhang, F. Lu, M. Xu, S. Liu, P.-C. Peng, S. Shen, et al., "An ultra-reliable MMW/FSO A-RoF system based on coordinated mapping and combining technique for 5G and beyond mobile fronthaul," Journal of Lightwave Technology, vol. 36, pp. 4952-4959, 2018.
- [85] G. Böcherer, F. Steiner, and P. Schulte, "Bandwidth efficient and rate-matched low-density parity-check coded modulation," IEEE Transactions on communications, vol. 63, pp. 4651-4665, 2015.
- [86] J. Cho, X. Chen, S. Chandrasekhar, and P. Winzer, "On line rates, information rates, and spectral efficiencies in probabilistically shaped QAM systems," Optics Express, vol. 26, pp. 9784-9791, 2018.
- [87] J. Cho, L. Schmalen, and P. J. Winzer, "Normalized Generalized Mutual Information as a Forward Error Correction Threshold for Probabilistically Shaped QAM," in 2017 European Conference on Optical Communication (ECOC), 2017, pp. 1-3.
- [88] S. Zhang and F. Yaman, "Design and comparison of advanced modulation formats based on generalized mutual information," Journal of Lightwave Technology, vol. 36, pp. 416-423, 2017.
- [89] C. R. S. Fludger and T. Kupfer, "Transmitter Impairment Mitigation and Monitoring for High Baud-Rate, High Order Modulation Systems," in ECOC 2016; 42nd European Conference on Optical Communication, 2016, pp. 1-3.
- [90] A. Ng'Oma, C.-T. Lin, L.-Y. W. He, W.-J. Jiang, F. Annunziata, J. Chen, et al., "31 Gbps RoF system employing adaptive bit-loading OFDM modulation at 60 GHz," in 2011 Optical Fiber Communication Conference and Exposition and the National Fiber Optic Engineers Conference, 2011, pp. 1-3.
- [91] L. Zhao, Y. Zhang, and W. Zhou, "Probabilistically shaped 64QAM OFDM signal transmission in a heterodyne coherent detection system," Optics Communications, vol. 434, pp. 175-179, 2019.
- [92] F. Steiner, P. Schulte, and G. Bocherer, "Approaching waterfilling capacity of parallel channels by higher order modulation and probabilistic amplitude shaping," in 2018 52nd Annual Conference on Information Sciences and Systems (CISS), 2018, pp. 1-6.
- [93] X. Hong, C. Fei, G. Zhang, and S. He, "Probabilistically shaped 256-QAM-OFDM transmission in underwater wireless optical communication system," in 2019 Optical Fiber Communications Conference and Exhibition (OFC), 2019, pp. 1-3.

- [94] C. Xie, Z. Chen, S. Fu, W. Liu, Z. He, L. Deng, et al., "Achievable information rate enhancement of visible light communication using probabilistically shaped OFDM modulation," *Optics express*, vol. 26, pp. 367-375, 2018.
- [95] S. Hu, W. Zhang, X. Yi, Z. Li, F. Li, X. Huang, et al., "MAP detection of probabilistically shaped constellations in optical fiber transmissions," in *Optical Fiber Communication Conference*, 2019, p. W1D. 3.
- [96] J. R. Barry, D. G. Messerschmitt, and E. A. Lee, *Digital Communication: Third Edition*: Kluwer Academic Publishers, 2003.
- [97] T. Mizuochi, "Recent progress in forward error correction and its interplay with transmission impairments," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 12, pp. 544-554, 2006.
- [98] C. Blair, "Problem complexity and method efficiency in optimization (as nemirovsky and db yudin)," *SIAM Review*, vol. 27, p. 264, 1985.
- [99] J. Zhang, K. Wang, Y. Wei, L. Zhao, W. Zhou, J. Xiao, et al., "280 Gb/s IM/DD PS-PAM-8 transmission over 10 km SSMF at O-band for optical interconnects," in *Optical Fiber Communication Conference*, 2020, p. M4F. 1.
- [100] D. Che and W. Shieh, "Achievable rate comparison between entropy and bit loading in a 100-Gb/s DM-DD DMT system," in *2019 Optical Fiber Communications Conference and Exhibition (OFC)*, 2019, pp. 1-3.
- [101] R. A. Shafik, M. S. Rahman, and A. R. Islam, "On the Extended Relationships Among EVM, BER and SNR as Performance Metrics," in *2006 International Conference on Electrical and Computer Engineering*, 2006, pp. 408-411.
- [102] "IEEE 802.3ca, standard for Ethernet - Amendment: Physical layer specifications and management parameters for 25 Gb/s and 50 Gb/s passive optical networks," June 2020
- [103] R. Kudo, T. Kobayashi, K. Ishihara, Y. Takatori, A. Sano, and Y. Miyamoto, "Coherent Optical Single Carrier Transmission Using Overlap Frequency Domain Equalization for Long-Haul Optical Systems," *Journal of Lightwave Technology*, vol. 27, pp. 3721-3728, 2009.
- [104] H. Yagi, T. Kaneko, N. Kono, Y. Yoneda, K. Uesaka, M. Ekawa, et al., "InP-based monolithically integrated photonic devices for digital coherent transmission," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 24, pp. 1-11, 2017.
- [105] Q. Zhuge, M. Morsy-Osman, X. Xu, M. E. Mousa-Pasandi, M. Chagnon, Z. A. El-Sahn, et al., "Pilot-aided carrier phase recovery for M-QAM using superscalar parallelization based PLL," *Optics Express*, vol. 20, pp. 19599-19609, 2012/08/13 2012.

- [106] E. Börjeson, C. Fougstedt, and P. Larsson-Edefors, "ASIC design exploration of phase recovery algorithms for M-QAM fiber-optic systems," in Optical Fiber Communication Conference, 2019, p. W3H. 7.
- [107] K. Sugihara, Y. Miyata, T. Sugihara, K. Kubo, H. Yoshida, W. Matsumoto, et al., "A spatially-coupled type LDPC Code with an NCG of 12 dB for optical transmission beyond 100 Gb/s," in 2013 Optical Fiber Communication Conference and Exposition and the National Fiber Optic Engineers Conference (OFC/NFOEC), 2013, pp. 1-3.
- [108] T. Pfau, "Carrier recovery algorithms and real-time DSP implementation for coherent receivers," in OFC 2014, 2014, pp. 1-17.
- [109] X. Wang, Y. Zeng, R. Liao, C. Zhao, H. Wu, and M. Tang, "Mach-Zehnder Interferometer based Endlessly Adaptive Polarization Controller on Silicon-Photonic Platform," in 2021 Optical Fiber Communications Conference and Exhibition (OFC), 2021, pp. 1-3.
- [110] T. Gui, J. Cao, X. Chen, K. Zheng, S. Yuan, X. Fang, et al., "Real-time Single-Carrier 800Gb/s DP-64QAM Demonstration using Bi-Directional Self-homodyne Coherent Transceivers with 200krad/s Endless Active Polarization Controller," in Optoelectronics and Communications Conference, 2021, p. T5A. 5.
- [111] L. Wang, Y. Chen, X. Wang, C. Zhao, J. Chen, and M. Tang, "Simplest DSP in Self-Homodyne Coherent Transmission Using Bidirectional Adaptive Polarization Controller," in 2021 Optical Fiber Communications Conference and Exhibition (OFC), 2021, pp. 1-3.
- [112] Y.-W. Chen, K. Kuzmin, M. Poirier, T. Tomimoto, G. Zarris, R. Moore, et al., "Over 60GHz InP CDM and ICR Enabling 800Gbps LR/ER/ZR/ZR+ Transmission Links With 120Gbaud/DP-16QAM Modulation," in 26th Optoelectronics and Communications Conference, Hong Kong, 2021, p. T5A.9.
- [113] K.-S. Kim, J.-H. Lee, W.-Z. Chung, and S.-C. Kim, "An electronic domain chromatic dispersion monitoring scheme insensitive to OSNR using kurtosis," Journal of the Optical Society of Korea, vol. 12, pp. 249-254, 2008.
- [114] P. J. Bickel, Y. a. Ritov, and A. B. Tsybakov, "Simultaneous analysis of Lasso and Dantzig selector," The Annals of Statistics, vol. 37, pp. 1705-1732, 28, 2009.
- [115] R. Tibshirani, "Regression shrinkage and selection via the lasso," Journal of the Royal Statistical Society: Series B (Methodological), vol. 58, pp. 267-288, 1996.
- [116] D. Chetverikov, Z. Liao, and V. Chernozhukov, "On cross-validated lasso in high dimensions," Annal. Stat.(Forthcoming), vol. 40, 2020.
- [117] A. Agarwal, S. Negahban, and M. J. Wainwright, "Fast global convergence of gradient methods for high-dimensional statistical recovery," The Annals of Statistics, pp. 2452-2482, 2012.

- [118] M. H. Gruber, "Statistical digital signal processing and modeling," ed: Taylor & Francis Group, 1997.
- [119] D. Berthelot, N. Carlini, I. Goodfellow, N. Papernot, A. Oliver, and C. Raffel, "Mixmatch: A holistic approach to semi-supervised learning," arXiv preprint arXiv:1905.02249, 2019.
- [120] S. Laine and T. Aila, "Temporal ensembling for semi-supervised learning," arXiv preprint arXiv:1610.02242, 2016.
- [121] D. G. Brennan, "Linear diversity combining techniques," Proceedings of the IEEE, vol. 91, pp. 331-356, 2003.
- [122] H. A. Mahmoud and H. Arslan, "Error vector magnitude to SNR conversion for nondata-aided receivers," IEEE Transactions on Wireless Communications, vol. 8, pp. 2694-2704, 2009.
- [123] Q. Yu, S. Corteselli, and J. Cho, "FPGA implementation of prefix-free code distribution matching for probabilistic constellation shaping," in Optical Fiber Communication Conference, 2020, p. Th1G. 7.

VITA

Rui Zhang was born in 1995 in Hefei, Anhui province, China. In 2017, she received her B.S. degree in Electrical engineering and B.A. in Economics both from Peking University, Beijing, China. Rui Zhang is currently a Ph.D. candidate at the School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, USA.

She joined Georgia Tech since fall 2017 and has worked on several projects for the National Science Foundation Center on Fiber-Wireless Integration and Networking. In 2018, 2019, and 2021, she was a research engineer intern in NeoPhotonics, San Jose, CA. In 2020, she was a research intern in Nokia Bell Labs, Murray Hill, NJ. Rui Zhang was the recipient of the several awards including Fall 2017 Oscar P. Cleaver Award from the Georgia Tech, 2018-2020 Corning Stem Scholarship, Bell Labs Summer Intern Award for Outstanding Innovation, and top scored paper in Optical Fiber Communication Conference 2022. Her research mainly focuses on joint hardware and signal processing design for next generation high-capacity communication systems.

She has authored and co-authored papers in prestigious IEEE and OSA journals and international conferences including 15 first-authored papers. She is an active reviewer for IEEE Communication Magazine, OSA Optics Express, OSA Optics Letters, IEEE/OSA Journal of Lightwave Technology, IEEE Photonics Technology Letters, OSA Journal of Optical Communications & Networking, IEEE Communication Letters, and IEEE Photonics Journal.