Datacenter Connectivity Technologies: Principles and Practice
In recent years, investments by cloud companies in mega data centers and associated network infrastructure has created a very active and dynamic segment in the optical components and modules market. Optical interconnect technologies at high speed play a critical role for the growth of mega data centers, which flood the networks with unprecedented amounts of data traffic.

Datacenter Connectivity Technologies: Principles and Practice provides a comprehensive and in-depth look at the development of various optical connectivity technologies which are making an impact on the building of data centers. The technologies span from short range connectivity, as low as 100 meters with multi-mode fiber (MMF) links inside data centers, to long distances of hundreds of kilometers with single-mode fiber (SMF) links between data centers. This book is the first of its kind to address various advanced technologies connecting data centers. It represents a collection of achievements and the latest developments from well-known industry experts and academic researchers active in this field.

Technological topics covered in this book include:
- Mega data center requirements
- High volume VCSELs
- Directly modulated lasers
- Electro-absorption modulated lasers
- Pulse amplitude modulation (PAM)
- Discrete Multi-Tone modulation (DMT)
- Optical Duobinary Transmission
- Optical fibers and connectors
- Mach-Zehnder modulators
- Silicon photonics
- Optical waveguide devices and packaging
- Testing and measurements
- Advanced modulation formats
- Optical coherent networks
- High-speed IC design & packaging

"Datacenter Connectivity Technologies: Principles and Practice, provides a comprehensive and in-depth look at the development of various optical connectivity technologies which are making an impact on the building of data centers. The technologies span from short range connectivity, as low as 100 meters with multi-mode fiber links inside data centers, to long distances of hundreds of kilometers with single-mode fiber links. The book is the first of its kind to address the various advanced technologies connecting data centers. It represents a collection of achievements and the latest developments from well-known industry experts and academic researchers active in this field."

"Datacenter Connectivity Technologies: Principles and Practice, provides a comprehensive and in-depth look at the development of various optical connectivity technologies which are making an impact on the building of data centers. The technologies span from short range connectivity, as low as 100 meters with multi-mode fiber links inside data centers, to long distances of hundreds of kilometers with single-mode fiber links. The book is the first of its kind to address the various advanced technologies connecting data centers. It represents a collection of achievements and the latest developments from well-known industry experts and academic researchers active in this field."
Demonstrations of CML3.6 New Era of High-Speed DML T
oward 100-GHz Bandwidth

3.6.1 Detuned-Loading Effect
3.6.2 S21 High-Pass Filter Effect Due to In-Cavity FM-AM
Conversion by the DBR Mirror
3.6.3 Photon-Photon Resonance Effect
3.6.4 Co-Existence of Photon-Photon Resonance and Detuned-Loading Effects
3.6.5 55-GHz Bandwidth Short-Cavity DR Laser and 56 Gbaud PAM4 Generation

3.7 Conclusions

References

4 PAM4 Modulation Using Electro-absorption Modulated Lasers

Trevor Chan and Winston Way

4.1 Introduction
4.2 General PAM4 Optical Transceiver and Link Considerations
4.2.1 PAM4 Signal and Optical Link Characteristics
4.2.2 EML Biasing and Nonlinear Equalization
4.2.3 Forward Error Correction (FEC) and Data Rates for PAM4 Links
4.2.4 Sampling Rate and Analog Bandwidth
4.2.5 FFE and DFE Equalization
4.3 28 Gbaud PAM4 Transmission
4.4 56 Gbaud PAM4 Transmission over 2 km Experiment
4.5 40 km PAM4 Transmission
4.5.1 Avalanche Photodiode (APD)
4.5.2 Gain Clamped Semiconductor Optical Amplifier (GC-SOA)
4.6 100 km PAM4 Transmission
4.6.1 Experimental Setup
4.6.2 Single Channel Characteristics
4.6.3 Effect of Fiber Nonlinearities
4.7 Multipath Interference
4.7.1 Experimental Demonstration of the Upper Bound MPI Scenario
4.7.2 Time-Domain Mixing Monte Carlo Simulation
4.7.3 MPI Experiment with Multiple Connectors
4.8 Summary

References

5 Optical Fiber for Datacenter Connectivity

Yi Sun and John Kamino

5.1 Introduction
5.2 Fiber Type for Datacenters
5.2.1 Multimode Fiber Types for Datacenters
5.2.2 Single-mode Fiber Types for Datacenters
5.2.3 Optical Cabling for Datacenters
5.2.4 Multicore (MCF) and Few-Mode Fiber (FMF) for SDM
5.3 Waveguide Design, Modal Structure, and Time Response of SMF and MMF for Datacenters
5.3.1 Fundamentals of Waveguide Design and Mode Structures of SMF and MMF
5.3.2 Fundamentals of the Time Response of Optical Fiber
5.4 Multimode Optical Fiber for High-Speed Short-Reach Interconnect
5.4.1 Laser-optimized MMF (OM3 and OM4)
5.4.1.1 What is Laser-optimized MMF?
5.4.1.2 Differential Modal Delay (DMD)
5.4.1.3 Bandwidth of MMF Links
5.4.2 Bend-optimized OM3/OM4 and Overfilled Effective Modal Bandwidth
5.4.3 Wideband MMF (OM5)
5.5 High-Speed VCSEL-MMF Short-Reach Optical Interconnect System
5.5.1 System Evaluation Methodology
5.5.2 High-Speed VCSEL-MMF System Transmission Validation
5.5.2.1 10GBASE-SR transmission over OM3 and OM4 MMF
5.5.2.2 40GBASE-eSR4 and 100G eSR4 extended reach demonstration over OM4 MMF
5.5.2.3 40/100 Gbps SWDM over OM5 MMF
5.5.2.4 High-Speed PAM4 SWDM transmission over OM5 MMF
5.6 Datacom Transmission over Single-Mode Optical Fiber

5.7 Conclusions

References

6 PAM4 Signaling and its Applications

Frank Chang

6.1 Introduction
6.2 A Brief History
6.3 PAM4 IC Implementation Challenges
6.3.1 PAM4 Transmit Architectures
6.3.2 PAM4 Receive Architectures
6.4 PAM4 SMF Performance
6.4.1 Experimental Setups
6.4.2 1λ 40G 10 km Transmission
6.4.3 2λ 100G 10 km and 40 km Transmissions
6.4.4 Technical Options for 200/400G Over SMF
6.5 PAM4 MMF Performance
6.5.1 Experimental Setups
6.5.2 1λ 40G Transmission Over 550m OM4
6.5.3 2λ 100/200 Gbps 300 m Transmission
6.5.4 Technical Options for 200/400G Over VCSEL/
Advanced Optical Measurements for Data Centers

Steve Yao, Wajih Daab, Gang He and Daniel Gariépy

12.1 Introduction

12.2 Polarization Related Tests

12.2.1 Polarization Mode Dispersion (PMD) Measurement

12.2.2 Polarization Dependent Loss (PDL) Measurement

12.2.3 PDR Measurement of Receivers

12.2.4 PDL Measurement of Fiber Optic Link

12.2.5 Measuring In-band OSNR by DOP Measurement

12.2.6 Polarization Emulation for Non-coherent and Coherent Systems

12.3 Optical Signal-to-Noise Ratio Measurement

12.3.1 Measuring OSNR with an OSA

12.4 Characterization of Optical Vector-Modulated Signals

12.4.1 Constellation and IQ Diagrams for Vector-Modulated Signal

12.4.2 Definitions of EVM, RMS-EVM and TR-EVM

12.4.3 Relationships between EVMRMS, Q-Factor, OSNR and BER

12.4.4 Characterization of Transmitter Impairments

12.5 Conclusion

References

Digital Signal Processing for Short-reach Optical Communications

Kangping Zhong, Xian Zhou, Jiahao Huo, Alan Pak Tao Lau, Chao Lu and Li Zeng

13.1 Introduction

13.1.1 Challenges for Short-reach Optical Systems

13.1.1.1 Cost

13.1.1.2 Form factor

13.1.1.3 Latency

13.1.2 Different Types of Short-reach Systems

13.1.2.1 Server-to-server or Intra-data-center links

13.1.2.2 Inter data-center links

13.1.2.3 Extended Reach Inter-data-center, Access, and Metro Links

13.2 Modulation Formats for Short-reach Systems

13.2.1 Pulse Amplitude Modulation (PAM)

13.2.2 Carry-less Amplitude and Phase (CAP) Modulation

13.2.3 Discrete Multi-tone (DMT) Modulation

13.2.4 Performance Comparison of Modulation Formats

13.2.5 Complexity Comparison of Modulation Formats

13.2.6 Recent Experiment on High-Speed Short-reach Transmission Systems

13.3 Digital Signal Processing for Short-reach Systems

13.3.1 Feed-forward Equalizer (FFE)

13.3.2 Decision Feedback Equalizer (DFE)

13.3.3 Direct Detection Faster-than Nyquist (DD-FTN)

13.3.4 Volterra-series Based Nonlinear Equalizer (VNLE)

13.4 Polarization Division Multiplexed Transmission for Short-reach Systems

13.4.1 Stokes-vector Direct Detection (SV-DD) Receiver

13.4.2 2-Dimensional (2D) PDM-DD System Based on SV-DD Receiver

13.4.3 3-Dimensional (3D) PDM-DD System Based on SV-DD Receiver

13.5 Conclusion

References

Multi-dimensional Polarization Modulation

Di Che, An Li, Xi Chen, Qian Hu and William Shieh

14.1 Optical Detection with Polarization Diversity

14.1.1 The Need of Polarization-Diversity Detection

14.1.2 Automatic Polarization Control

14.1.3 Polarization-Diversity Detection in Jones Space

14.1.4 The Barrier of Self-Polarization Diversity

14.1.5 Polarization-Diversity Detection in Stokes Space

14.2 Direct Modulation with Coherent Receiver

14.2.1 The Intensity-only POL-MUX-DM Coherent System

14.2.2 Complex DM Model

14.2.3 100-Gb/s CDM Transmission Over 1600-km SMF

14.3 Polarization Modulation in Stokes Space

14.3.1 Stokes-space Modulation

14.3.2 Universal MIMO Equalization in Stokes Space

14.3.3 Self-coherent SSM

14.3.4 Multi-Dimensional IM in Stokes Space

14.4 Noncoherent Polarization Multiplexing

14.4.1 Degree of Coherence in POL-MUX Transmitter

14.4.2 Noncoherent POL-MUX Schemes

14.5 Summary

References

High-speed Flexible Coherent Optical Transport Network

Tiejun J. Xia and Glenn A. Wellbrock

15.1 Introduction

15.2 Why Optical Coherent Transmission?

15.3 Optical Transport Network with...
Coherent Transmission

15.4 What's Next for Optical Transport Network?

15.5 Coherent Transport Technology Development by Network Operators

15.6 Datacenter Connections and Coherent Transport Networks

15.7 Conclusions

References

16 Ultra-low-power SiGe Driver-IC for High-speed Electro-absorption Modulated DFB Lasers

Jung Han Choi

16.1 Introduction

16.2 IC Design for Low Power Consumption

16.2.1 Design Requirements

16.2.2 IC Architectures for Low Power Consumption

16.2.3 Driver-IC Design

16.2.3.1 Unit-cell design

16.2.3.2 Circuit Simulations

16.3 Co-design and Electro-optical Simulation

16.3.1 Low-power CoC Design

16.3.2 Co-simulation of Driver-IC with EML

16.3.2.1 Electrical Time-domain Simulations

16.3.2.2 Electro-optical Time-domain Simulations

16.4 Measurements

16.4.1 EO Measurements

16.4.2 Transmission Experiments

16.5 Conclusion and Perspective

References

Index

About the Editor

Preface

Welcome to "Datacenter Connectivity Technologies: Principles and Practice". This book discusses relevant concepts and inherent technologies that should be taken into account when designing and implementing data center connectivity solutions. It was motivated by the desire to explore the future perspectives of high-speed data center connectivity technologies by presenting the state-of-the-art results in both new optical devices, and circuit implementations. It's well known that, in recent years, investments by cloud companies in mega data centers and associated network infrastructure has created a very active and dynamic segment in the optical components and modules market. Optical interconnect technologies at high speed play a critical role for the growth of mega or hyperscale data centers, which flood the networks with unprecedented amount of data traffic. And yet, there is only a limited number of books in this area that give a coherent and comprehensive review of data center technologies in general. Most of the data center related research materials are scattered around in journals, periodicals, conference proceedings, and a number of technical standards by industrial forums. Therefore we feel it is timely to publish a book that covers the different aspects of the data center connectivity technology. One may reliably count on this book as the first of its kind to address various advanced technologies connecting data centers.

This book provides a comprehensive and in-depth look at the development of various optical connectivity technologies which are making an impact on the building of data centers. The technologies span from short range connectivity, as low as 100 meters with multi-mode fiber (MMF) links inside data centers, to long distances of hundreds of kilometers with single-mode fiber (SMF) links between data centers. It is one of the purposes of this book to provide a balanced coverage of networking technologies, fiber optics transmission technologies, and electronics/components involved in developing data center connectivity solutions.

One point I'd like to mention is that electronic power is the lifeline of the data center. Since power represents up to 70% of the total operating costs, many enterprise users and colocation operators focus on their technology and site selection on low-cost-power options. Very often, researchers in new technology such as data centers will be driven by the desire to demonstrate technical smartness and overlook practicality. This tends to result in a lot of literatures which eventually becomes irrelevant due to economic reasons. Data center
connectivity needs to be very cost-conscious with low power consumption. Therefore, when selecting the materials for this book, we tried to balance research interest with practical economic and engineering considerations. This book represents a collection of recent achievements and the latest developments from well-known industry experts and academic researchers active in this particular field. The book is organized in 16 chapters, covering different aspects of data center connectivity technologies. Chapter 1 begins with a contribution by Alibaba distinguished researchers and discusses from its networking perspective the requirements and technologies for the optical interconnect technologies in datacenters, including intra-datacenter and inter-datacenter interconnects. Alibaba is the world's largest retailer with online sales and profits surpassing Walmart, Ebook Tops, and eBay combined. Alibaba is expanding its cloud data center footprint from 14 data centers in operation and growing rapidly to support its global expansion. Chapter 2 is written by semiconductor laser veteran who pioneered the Vertical cavity surface emitting lasers (VCSEL) from its early days of fabrication. VCSELs have been the primary laser sources for short-distance datacom and data centers over multimode fibers from 1 to 25 Gbps per channel in the form of either serial or parallel optical data links. Chapter 3 and 4 review two critical transmitter sources: directly modulated laser diodes (DML) and electro-absorption modulated lasers (EML) from two industry leading suppliers by Finisar and NeoPhotonics separately. DML and EML are widely used by default for data centers and client side transceivers. Chapter 5 provides an overview of optical fibers and connectors used for data center connectivity from the perspective of fiber manufacturer. Multimode fibers still has many advantages in short-reach applications that have been enhanced by the introduction of WDM technology. On the other hand, single-mode fibers have become more prevalent in hyper-scale cloud based datacenters that have much longer reach requirements. Chapter 6 to 8 discusses various optical modulation technologies for direct detection: PAM4, DMT and Duobinary. PAM4 has been adopted by IEEE802.3 as signaling standard by default for Ethernet inside data centers. DMT was introduced to boost the line rate with the limited bandwidth of optoelectronic devices. It may play an important role in Metro DCI. Three-level Duobinary may also be useful for power-efficient ultra-high serial rate optical links. Chapter 9 describes the basic principles and presents industrial features of LiNbO3 (LN) Mach-Zehnder Modulator (MZM) which has been successfully operated in long distance and high capacity optical fiber transmission systems for over 30 years. Compact size with higher bandwidth LN-MZM is strongly demanded especially in the field of application related to the data center. New technologies are expected to be introduced in the production of next generation LN-MZM. Chapter 10 provides an overview of implementing the PAM4 and Si Photonics for achieving the 100 Gbit/s, DWDM datacenter interconnections that span regional distances of up to 120 km. The combination of Si photonics for the highly integrated optical components, and high speed Si CMOS for signal processing is critical for the implementation of low-cost, low-power, switch pluggable optical modules. Chapter 11 is devoted to the ultra-low-loss photonics integrated circuits (PIC) and packaging, contributed by researchers of UC Santa Barbara. PIC is a device that integrates multiple (at least two)
photonic functions in optical domain and as such is similar to an electronics integrated circuit. Compact PICs are critical to address cost, size, weight, and form factor. One PIC example is integrated dispersion compensation fabricated in a low loss silicon nitride platform mitigating PAM4 dispersion for multiple WDM channels while at the same time satisfying the strict OSNR requirements. Chapter 12 is instructive in terms of the common measurement techniques used for data centers, especially for high speed optical measurements at longer distance. The authors discuss in details the polarization effects, OSNR measurements and characterization of optical vector modulated signals such as what's used in coherent detection. Chapter 13 provide a concise summary of various digital signal processing (DSP) technique, while Chapter 14 describes in details the multi-dimensional polarization modulation. The research community has put tremendous efforts to mirror the achievements in long-haul coherent transceivers for developing DSP strategies suitable for short-reach datacenter connectivity. If readers are interested in designing solutions beyond 400Gb/s, it is recommended to read through those two chapters written by DSP experts and testing professionals. Chapter 15 is contributed from network operator perspective to look into the high speed flexible coherent optical transport networks. Flexible coherent transport supports multiple baud rates, modulation orders, and data rates to serve more capacity for inter-datacenter applications. Coherent tends to move to shorter reaches as baud rate increases, for example 400 Gbit/s switch pluggable 16QAM coherent modules are being standardized to address next-gen DCI modules. Chapter 16 is devoted to present ultra-low-power SiGe Driver IC for high-speed EMLs. It's emphasized that co-design of the driver IC with the EML could enhance the TOSA power efficiency at high baud rates. If readers are interested in designing analog ICs working at beyond 53Gbaud rates, it is recommended to read through this chapter written by SiGe device expert. It is our wish to present this book as a comprehensive reference of data center connectivity technologies for those interested in developing and understanding this fast-growing area. I sincerely believe that it will provide a major boost in your understanding of the various latest technologies after reading through the chapters by the international experts in industry and academia. The discussion points range from system requirements to component/IC design. They are useful to understand current state-of-the-art technologies in these development fields. Also relevant standard activities and technical details behind those are addressed. The readers can grasp the future potential of these interconnect technologies and research trends by reading this book. This book is intended as a general reference for researchers, senior and graduate-level college students working in the field of data center networks. It can be also used by engineer and managers to obtain a working knowledge of data center connectivity technologies. The book provide the breadth for people who need to gain a generic understanding of this specific field, and the depth, for those who would like to dig deeper into data center connectivity technologies, and relevant areas. Frank Chang, Ph.D Silicon Valley, 2018 fymchang@gmail.com

Acknowledgements First, I would really like to thank all the authors who have made tremendous efforts to contribute to this book out of their extreme busy schedule.
I will always remember their contribution and active exchange of their thoughts and ideas. I would like to thank River Publishers and Rajeev Prasad for giving me the opportunity to start this project and edit this comprehensive and valuable book. Special thanks to Junko Nakajima, our production manager, who has been extremely patient with my constant slipping of the schedule, and also to her for her excellent project management. Last but not the least, I want to express my gratitude to my wife, Jenny. She was the first to provide invaluable insights to my work in progress during my completing this work. Since this book was completed mostly in my spare time at home with endless weekends, she gave up a lot of her free time to support me working on this project.

Frank Chang

List of Contributors

Alan Pak Tao Lau, Department of Electrical Engineering, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong SAR, China

An Li, Futurewei Technologies, Santa Clara, CA, USA

Chao Lu, Department of Electronic and Information Engineering, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong SAR, China

Chongjin Xie, Alibaba Infrastructure Service, Alibaba Group, Sunnyvale, CA, USA

Daniel J. Blumenthal, University of California Santa Barbara, California, USA

Daniel Gariépy, EXFO Inc., Quebec, Canada

Demis D. John, University of California Santa Barbara, California, USA

Di Che, The University of Melbourne, Melbourne, Australia

Frank Chang, Inphi Corp, California, USA

Gang He, EXFO Inc., Quebec, Canada

Grant Brodnik, University of California Santa Barbara, California, USA

Glenn A. Wellbrock, Verizon, Richardson, TX, USA

Gordon Ning Liu, Huawei Technologies Co., Ltd., Shenzhen, China

Guy T orfs, imec - Ghent University, IDLab, Gent, Belgium

Hirochika Nakajima, Waseda University, Tokyo, Japan

Jiahao Huo, University of Science and Technology Beijing, Beijing, China

Johan Bauwelinck, imec - Ghent University, IDLab, Gent, Belgium

John Kamino, OFS Fitel LLC, Norcross, GA, USA

Jung Han Choi, Fraunhofer Heinrich-Hertz-Institute, Berlin, Germany

Kangping Zhong, MACOM Technology Solutions Inc, Shenzhen, China

Li Zeng, Huawei Technology Ltd, Shenzhen, China

Liang Zhang, Huawei Technologies Co., Ltd., Shenzhen, China

Mark Filer, Microsoft Corporation, Redmond, WA, USA

Michael Belt, Honeywell Inc., California, USA

Qian Hu, Nokia Bell Labs, Stuttgart, Germany

Radhakrishnan Nagarajan, Inphi Corporation, Santa Clara, CA, USA

Renan L. Moreira, University of California Santa Barbara, California, USA

Sarat Gundavarapu, University of California Santa Barbara, California, USA

Steve Yao, General Photonics Corporation, California, USA

T aran Huffman, GenXComm Inc., Austin, TX, USA

T rervention Chan, Neophotonics, California, USA

Tiejun J. Xia, Verizon, Richardson, TX, USA

Wajih Daab, General Photonics Corporation, California, USA

Wenbin Jiang, WJ Technologies LLC, California, USA

William Shieh, The University of Melbourne, Melbourne, Australia

Winston Way, Neophotonics, California, USA

Xi Chen, Nokia Bell Labs, Holmdel, New Jersey, USA

Xian Zhou, University of Science and Technology Beijing, Beijing, China

Xin (Scott) Yin, imec - Ghent University, IDLab, Gent, Belgium

Y asuhiro Matsui, Finisar Corporation, Fremont, CA, USA

Yi Sun, OFS Fitel LLC, Norcross, GA, USA

Yuya Y amaguchi, National Institute of Information and Communications Technology (NICT), Tokyo, Japan

List of Figures

Figure 1.1 Global datacenter internet protocol
feedback phase conditions.

Figure 3.48 Round trip gain (blue curve) on the left and phase (red dashed curve) on the right at the DBR laser threshold. The squared red marker represents the lasing mode; the blue markers indicate non-lasing cavity modes. The green asterisks on the reflectivity curve represent the modes locations in the maximum detuned loading condition.

Figure 3.49 Schematic picture of a two section DBR laser with GSG contact configuration, (b) Small signal modulation response of a DBR laser lasing on the long (right) and short (left) wavelength sides of the Bragg reflection peak.

Figure 3.50 (a) Schematic side view of two-section DR laser where the light is coupled through DBR transmission, and (b) corresponding DML structure with PN blocking with double-channel stripe filled with BCB.

Figure 3.51 (a) S21 response of DR laser for a bias range from 10 mA to 36.2 mA. Subtracted S21 response between 36.2 mA and 30 mA is shown in red broken line, (b) AM and FM response at 36.2 mA bias.

Figure 3.52 (a) RIN spectra for another DR laser showing Fr and PPR peaks both around 40 GHz at 45 mA, and (b) corresponding S21 response.

Figure 3.53 (a) 56-Gb/s NRZ, (b) 112-Gb/s PAM4 back-to-back, (c) transmission with 7 ps/nm fiber dispersion (∼3 km at 1330 nm), and (d) –28 ps/nm fiber dispersion, (e) BER for BtB, positive (7 ps/nm) and negative dispersion (–28 ps/nm). The inset is the eye after non-linear equalization.

Figure 4.1 BER vs. SNR relationship for PAM4.

Figure 4.2 Theoretical BER vs. Received optical power for 28 GBaud and 56 Gbaud PAM4 transmission using state-of-the-art components.

Figure 4.3 Electrical to optical relationship of an EML. With PAM4 signals, the signal can be a) limited to the linear region, b) have increased amplitude, thereby inheriting nonlinearities of the transfer curve or c) pre-compressed so the output signal has equal eye openings.

Figure 4.4 a) A PAM4 signal with bandwidth in excess (2x) of the Nyquist frequency and b) a PAM4 signal bandwidth near the Nyquist frequency showing nearly closed eyes.

Figure 4.5 Reduced OMA after applying pre-emphasis.

Figure 4.6 Basic setup of the 28 Gbaud PAM4 link. Experimental and simulated results of two different bandwidth EMLs. Sensitivity and error floor improvement is gained using higher bandwidth EMLs with proper level equalization.

Figure 4.7 (a) Basic setup for a 56 Gbaud PAM4 system. The pre-distorted electrical and optical eye diagrams show all three eyes are open. (b) The frequency response of the combination of DAC, driver, and EML without (blue) and with (green) pre-distortion (FFE taps shown in the inset). Generally speaking EML's bandwidth still needs further improvement, and it is the main limiting factor for today's 53 Gbaud EML's TDECQ.

Figure 4.8 Results using the 56 Gbaud PAM4 link.

Figure 4.9 Similar to Figure 4.8 but with an improved receiver performance. 56 GBaud BER vs. Outer OMA when the signal is transmitted through 0 and 10 km of fiber.

Figure 4.10 28 Gbaud PAM4 transmission using an APD or a GC-SOA for a 1308 nm signal wavelength.

Figure 4.11 Experimental Setup for two 28 Gbaud PAM4 wavelengths passing a GC-SOA after 40 km SMF.

Figure 4.12 28 Gbaud PAM4 BER vs. received optical power for systems using a gain clamped SOA and PIN as the receiver.

Figure 4.13 Experimental setup using an GC-SOA to transmit a 56 Gbaud PAM4 signal over 40 km SMF.

Figure 4.14 56 Gbaud PAM4 BER vs. power received into the SOA using different gain clamp power levels. Back-to-back and after 40 km SMF with yes and no equalization.
km SMF transmissions are shown. With a –3 dBm gain clamp, a 1.5 dB margin after 40 km SMF transmission can be obtained.

Figure 4.15 System for 100 km transmission using a 10 G, C-band EML.

Figure 4.16 100 km system measurement of (a) fiber dispersion sensitivity and (b) BER vs. OSNR.

Figure 4.17 BER vs. fiber launch power under different received optical power.

Figure 4.18 A diagram illustrating multipath interferences due to doubly reflected light paths: (I) between connectors; (II) between a connector and TX reflectance; and (III) between a connector and RX reflectance.

Figure 4.19 MPI interferometer used to introduce multipath interference into the signal. The bottom arm in the interferometer is the path for MPI whose polarization is aligned to the signal for maximum interference.

Figure 4.20 BER vs. inner OMA for various MPI powers.

Figure 4.21 Monte Carlo histogram of MPI crosstalk powers from randomized polarization, phase, and symbol levels from each connector reflection in a triple link system. Sample size = 40,000.

Figure 4.22 Experimental setup used to create randomized MPI in a PAM4, a) triple link channel, b) the experimental setup based on a triple link model where the connector reflectance was purposely polished to be –26 dB, and c) the Poincare sphere showing significant polarization variation by shaking the fiber over 8 hours. When the fan was off, the SOP was spread in the narrow red arc range.

Figure 4.23 Results from both the Monte Carlo simulation and laboratory triple link experiment using connectors with –26 dB return loss. The experimental sample size can be compared with the simulations to extrapolate the long term MPI penalties.

Figure 4.24 Results from both the Monte Carlo simulation and laboratory triple link experiment using connectors with –35 dB reflectivity. Large MPI sample sizes do not show much increase in power penalty due to the low MPI power from the connectors.

Figure 5.1 Illustration of a global optical network formed by datacenters and inter-datacenter backbone optical cables.

Figure 5.2 Physical structure and refractive index profiles of optical fibers. (a) Side view of an optical fiber; (b) cross section view of an optical fiber; (c) refractive index profile of a graded-index fiber; (d) comparison between the index profiles of MMF and SMF.

Figure 5.3 Normalized propagation constant $\beta_l$ versus V-number for a step-indexed fiber (colored lines) and for 50/125 µm GI-MMF at 850 nm.

Figure 5.4 Effective index of LP modes at 850 nm for 19 PMG of 50/125 µm GI-MMF.

Figure 5.5 Radial mode power distribution of the ten LP modes in PMG19.

Figure 5.6 Index versus wavelength from 750 to 1750 nm for pure silica, 13.5 mole% Ge-doped Silica, and 1 mole% F-doped Silica.

Figure 5.7 Group delay versus wavelength from 750 to 1750 nm for pure silica, 13.5 mole% Ge-doped silica, and 1 mole% F-doped silica.

Figure 5.8 Chromatic dispersion versus wavelength for pure silica, 13.5 mole% Ge-doped silica, and 1 mole% F-doped silica.

Figure 5.9 Group delay and chromatic dispersion of LP01, LP11, LP02, and LP91 modes for GI-MMF optimized at 850 nm.

Figure 5.10 Top: delays of all LP modes versus effective index difference. Bottom: Modal delay of PMG versus principal mode number from 850 to 1300 nm for GI-MMF.

Figure 5.11 DMD scan for an OM3-grade MMF and DMD mask sets.

Figure 5.12 OM4 DMD traces (blue) and DMD masks. Measured DMD traces at 850 nm for an OM4 fiber. The 0–18 µm, 0–23 µm, and sliding mask widths are 0.043, 0.049, and 0.037.
back and over 25 km, 35 km and 40 km of SMF-28 SSMF with the KP4-FEC 2.4E-4 threshold indicated and (b) four 100 GHz spaced channels λ0 - λ3 with received optical power (ROP) = –7 dBm also showing performance well below the KP4-FEC 2.4E-4 threshold (channel λ was not recovered for 40 km due to total dispersion exceeding compensated + residual).

Figure 11.15 Schematic representation of a linear Bragg grating structure with periodic width perturbations along the waveguide propagation length. The period of the structure is denoted as Λ.

Figure 11.16 Measured and simulated (FIMMWAVE and T-matrix formalism grating bandwidth for 1000 μm long gratings under TE or TM excitation. The plot also gives fitted and simulated coupling constant values for the same set of gratings. The nominal waveguide width is 2.8μm and the waveguide core thickness is 80 nm. For an increased core thickness it is expected that these measured κ values would decrease. This is because as the core thickness increases, a changing waveguide width has less impact on an overall effective index change for the structure.

Figure 11.17 Measured grating coupling constants for 1 cm long gratings under either TE or TM excitation. The nominal waveguide width is 2.8 μm and the waveguide core thickness is 80 nm. The plot also gives fitted coupling constant values (κ) for the same set of gratings.

Figure 11.18 Illustration of wavelength selectivity of an add/drop ring resonator.

Figure 11.19 Ring resonator spectra at 1550 nm for R = 9.8 mm, α = 5 dB/m, n = 1.48. (a) Add-drop spectra highlighting the FSR and FWHM, and varying kc for (b) single bus ring, (c) drop port of 2 bus ring (kc1 = kc2), and (c) through port of 2 bus ring (kc1 = kc2).

Figure 11.20 (a) Third-order ring filter design; (b) Schematic cross section of Si3N4 low loss waveguide. In this device, we use nitride core thickness t1 = 175 nm, core width w = 2.2 µm and upper cladding thickness t2 = 6.8 µm. Thermal oxide lower cladding thickness is 15 microns.

Figure 11.21 (a) Photo micrograph of the third-order filter. (b) Image of a 3.5 mm wide bar of 5 third-order filters relative to a quarter.

Figure 11.22 Schematic representation of measurement setup.

Figure 11.23 Wavelength sweeps of third-order filters. The measurement is limited by photodetector dynamic range. (a) shows a filter initially out of resonance, (b) shows the same filter tuned to resonance.

Figure 11.24 (a) Third-order filter function, with an extinction ratio of 80 dB and FSR 48.2 GHz. The analytical fit yields κ1 fit = 0.125, κ2 fit = 0.005. (b) Third-order filter passband with a shape factor of 0.437 and no ripple.

Figure 11.25 A third-order ring filter is tuned over its full FSR. Tune 1 represents no thermal tuning, tune 2 represents 50 mW of thermal tuning, and tune 3 represents 110 mW of tuning.

Figure 11.26 First order rings with and without a metal layer are compared. Fitting the two curves to the theoretical model yields an additional loss, due to the metal layer of, 1.7 dB/m.

Figure 11.27 (a) Cross-sectional geometry of an on-chip Si3N4-core/SiO2/Al2O3-clad rare-earth-ion-doped waveguide amplifier. The Al2O3 layer will act as a host for any dopant ions incorporated during the deposition process. The respective layer thicknesses are as follows: t1 = 15.0 µm, t2 = 0.08 µm, t3 = 0.1 µm, t4 = 1.5 µm. The width of the Si3N4-core is 2.8 µm. It is this Si3N4-core that provides the majority of the lateral guiding of the optical mode. (b) Three-dimensional schematic image of the geometry. (c) Simulated (via FIMMWAVE) optical mode profile of the fundamental TE waveguide mode at the 1.55 µm
In 1965, Kao with Hockham found that by removing the impurity in glass material the fundamental limitation for light attenuation in glass is below 20 dB/km [1]. In 1970, Schultz, Keck, and Maurer in Corning made the first low-loss optical fibers, with a loss coefficient of 17 dB/km [2]. These started the era of fiber-optic communications. The first field experiment of fiber-optic communication systems was conducted in 1976, at a bit rate of 45 Mbps [3]. Since then, significant progress has been made in the field. With the advances in technologies such as time-division multiplexing (TDM), wavelength-division multiplexing (WDM), polarization-division multiplexing, digital coherent detection, etc., the capacity of fiber has increased by more than 10^6 times. Single-mode fiber (SMF) capacity has reached 100 Tbps [4], and by using multimode and multi-core technologies, over 2-Pbps capacity on a single fiber has been achieved [5, 6].

In the past, optical communications had mostly been used by telecom...
operators in their long-haul, metro, and access networks, to connect their central offices with long-haul and metro networks and bring their customers to central offices with access networks.

In 2008, the demand for optical communications from hyper-scale datacenters exceeded that of telecom operators in the United States [7]. Today optical interconnect demand from datacenters has become a major driving force for optical communications. Internet services have become part of our daily lives. Almost all the internet services such as web-browsing, e-mail, e-commerce, video streaming, social networking, and cloud computing run in datacenters, where a massive number of servers are connected and work together like a supercomputer [8], which generates a huge amount of internet traffic both inside and outside of datacenters. Most internet traffic has originated or terminated in a datacenter since 2008, and datacenter traffic will continue to dominate internet traffic in the foreseeable future. As shown in Figure 1.1, the annual global datacenter internet protocol (IP) traffic will reach 15.3 ZB (Zettabytes, Zetta = 10^21) in 2020, up from 4.7 ZB in 2015, a three-fold increase in 5 years with a compound annual growth rate (CAGR) of 27% in the 5-year time frame, driven mainly by video streaming and cloud computing [9].

The distribution of global internet traffic by destinations forecasted by Cisco is given in Figure 1.2 [9], which shows that by 2020 non-datacenter-related internet traffic only accounts for less than 1% of the total IP traffic, and more than 99% of internet traffic will be originated and/or terminated in a datacenter. Datacenters have replaced telecom center offices to become the center of internet traffic.

Figure 1.1 Global datacenter internet protocol (IP) traffic growth.

A typical architecture of datacenter networks is shown in Figure 1.3. In general, a worldwide hyper-scale datacenter operator has datacenters distributed around the world. They divide the world into many regions, and these regions are connected with optical mesh networks. Each region has one or a few gateways connecting to public internet, through which users can access their datacenters. In each region, there are one or a few metro areas, depending on customer distributions. Typically, a few datacenters are set up in a metro area with distances among them limited mostly to 80 km, which is the consideration of both disaster prevention and the latency requirement set by synchronous replication among different datacenters. To improve user experience and reduce the latency for customers connecting to their networks, many points of presence (POPs) may be built close to customers. The POPs have direct connections to datacenters so that they can quickly bring customer traffic to datacenters. These POPs can also be used as edging computing sites where data can be computed and processed locally. This brings two benefits: (1) the computation is close to data sources so that intelligence can be provided locally with reduced latency; (2) most data are processed locally so that the bandwidth requirement on the connections from POPs to datacenters can be significantly reduced. Inside datacenters, servers are connected through intra-datacenter networks.

Figure 1.2 Global internet protocol (IP) traffic by destination in 2020.

From Figure 1.3 and the above descriptions, one can see that a data-center network infrastructure includes access networks (connecting POPs to datacenters), backbone networks (connecting datacenters in different metro areas), metro networks (connecting datacenters...
within a metro area), and intra-datacenter networks (connecting servers inside a datacenter). Each of these networks has its own characteristics and requirements. The distance ranges for these networks are quite different. Intra-datacenter distances are less than 2 km and metro-network links are less than 80 km, but for backbone and access networks, hundreds to thousands of kilometers are needed. The different length requirements of different networks result in different optical interconnect technologies. In the following sections, we will discuss the details on optical technologies for intra-datacenter interconnects and inter-datacenter interconnects in metro networks and wide-area-networks (WANs).

Figure 1.3 Architecture of a datacenter network, including intra-datacenter networks, metro-networks, backbone networks, and access networks. POP: point of presence; DC: datacenter.

1.2 Intra-datacenter Interconnects

The computing and storage powers needed by internet services and cloud computing are provided by a massive number of servers located in each datacenter. Servers in each datacenter are connected with intra-datacenter networks so that they can communicate and exchange information with each other. The ideal solution is to use one big switch to connect all these servers, but this solution is technically difficult and cost prohibitive. In practice, an intra-datacenter network is built with many small switches and they are architected in multiple layers. Lots of network architectures have been proposed and implemented for datacenters, with the aim to reduce cost and increase scalability and efficiency for a given bisection bandwidth and number of connected servers [10, 11]. Figure 1.4 shows a typical architecture of an intra-datacenter network. It has three layers of switches. The edge switches are directly connected to servers, and the aggregate switches and core switches are used to increase the size and the scalability of the network. The incoming and outgoing traffic is handled by the routers and transmitted to other datacenters in metro networks or wide-area networks by optical transport networks. Due to the massive number of servers in a datacenter, there are lots of links in a datacenter network, as well as switches. For a datacenter with 10,000 servers, the numbers of links and switches can be from 15,000 to 30,000 and 1,000 to 3,000, respectively, depending on the network architecture [10].

Figure 1.4 Intra-datacenter network architecture. SW: switch.

At low speeds, these switches and servers are connected with copper cables. With the increase of network speeds and sizes, copper cables may not be able to meet most of the interconnect requirements. From 1 Gb-Ethernet, optical technologies start to be widely used in datacenters. In today's datacenters, optical interconnects are used in almost every connection outside of servers, providing high-bandwidth channels between connected objects. At the interface of an electrical switch, optical transceivers are used to convert the outgoing signals from the electrical domain to the optical domain and the incoming signals from the optical domain to the electrical domain, and optical fibers are used as media to transmit optical signals from one location to another, as shown in Figure 1.5 [12].

Figure 1.5 Diagram of optical transceivers in an electrical switch.

There are three different speeds in datacenter networks, server speed (network-interface-card speed of a server), switch speed, and router speed, and datacenter network speed usually refers to the...
switch speed. Datacenter traffic is mostly Ethernet traffic. Figure 1.6 gives the Ethernet speed evolution published by Ethernet Alliance, from 10G to Tb [13]. There are three speed evolution curves. One curve is for “Serial Speeds”, which are determined by the speeds of SERDES and are usually the same as the speeds of servers. One way to get to higher speeds is to use multiple parallel lanes, and four lanes are widely used for switches, as the “Quad Speeds” curve shows. For example, in 40G networks, server speed is 10G, and the switches use four 10G lanes to get to 40G. The next server speed after 10G is 25G and the corresponding switch speed is 100G, which can be supported by 25G serial speed and 100G quad speed that has already been standardized and deployed, as shown in Figure 1.6. The datacenter speeds after 100G is either 200G or 400G. 200G networks can be supported by 50G serial speed for servers and 200G quad speed for switches as illustrated in Figure 1.6. The best way to realize a 400G datacenter network is to use a 100G serial speed for servers and a 400G quad speed for switches, but due to the limitations of SERDES, 100G serial speed may not be realized until 2025 according to Ethernet Alliance Roadmap. Before that, 400G networks can be achieved by using higher parallelism, for example, 2 × 50G for 100G servers and 8 × 50G for 400G switches.

The “Highly Parallel Speeds” curve in Figure 1.6 is in general for routers and optical transport equipment, which typically demand higher speeds than switches and can be realized with more than four parallel lanes. For example, early-day 100G and 400G are realized with ten 10G lanes and sixteen 25G lanes, i.e., 10 × 10G and 16 × 25G, respectively.

Figure 1.6 Ethernet speed roadmap issued by Ethernet Alliance.

Figure 1.7 Historical view of front panel bandwidth of switch I/O and pluggable optics per RU. RU: rack unit. Courtesy of X. Zhou.

Due to short reaches and large quantities, there are different requirements on connections inside datacenters from those used in long-haul, metro, and access networks, so different technologies have been developed. Most links inside datacenters are in the range of a few meters to a few hundred meters. There are some long links between different buildings in a campus, with the longest link limited to 2 km. Bandwidth cost is the primary requirement for intra-datacenter connections. When calculating cost, the cost of both optical transceivers and fibers need to be included. The second requirement is bandwidth density or faceplate density, which can be quantified by the capacity that the front panel of a 19-in 1-RU (482.60mm × 44.45mm) switch can accommodate. The bandwidth of switch I/O still evolves in Moore’s law, as shown in Figure 1.7 [14]. To fully utilize the switch bandwidth, the bandwidth density of optical interconnects must be able to scale in the same speed. Figure 1.7 shows that before 2010, the bandwidth density of pluggable optical transceivers is higher than that of switch I/O, but it is not anymore after 2010. The bandwidth density is determined by not only the speeds, but the form factors of transceivers as well. Figure 1.8 compares the port density of different form factors, i.e., how many transceivers of different form factors can fit in a 1-RU face plate. There are more than a dozen optical transceiver form factors, and SFP and QSFP have become the main transceiver form factors in datacenters. The third requirement is the power consumption of optical transceivers. Although the power consumption of optical transceivers or even networks is a small portion of the total
power consumed by a datacenter, power-efficient optical transceivers are critical for high bandwidth-density switches as low power consumption optical transceivers can help power dissipation and increase bandwidth density. Figure 1.8 Port density comparison for different form factors.

1.2.1 40G Optical Interconnect Technologies

Up to 40G, most intra-datacenter optical interconnects use multimode technology, i.e., vertical-cavity surface-emitting lasers (VCSELs) combined with multimode fiber (MMF). VCSELs not only have a vertical structure so that thousands of VCSELs can be processed simultaneously on a single wafer, but are of low power consumption as well. Compared with SMF, MMF has a larger core size and numerical aperture so that it is much easier to couple light in and out MMF than SMF. Due to these two reasons, multimode technology is widely used in short-reach applications. However, light launched into MMF excites different modes, which travel at different speeds and limit the bandwidth of MMF, as shown in Figure 1.9. To increase the bandwidth of MMF, four generations of MMF have been developed, from OM1 to OM4. Table 1.1 gives the characteristics of different MMF. The bandwidths and link distances have been significantly improved from OM1 to OM4. For example, the link distance of a 10G system has been increased from 33 m for OM1 to 550 m for OM4. Note that the link distances listed in Table 1.1 are defined by IEEE standard [15, 16], and the link distances can be further increased using some distance-enhancement technologies.

Figure 1.9 Schematic of pulse propagation in multimode and single-mode fibers.

Table 1.1 Characteristics of different multimode fibers

There are mainly two types of 40G optical transceivers – one uses multi-mode technology, called SR4, and the other single-mode technology, called LR4. Both are standardized by IEEE [15]. The schematic of a 40G SR4 optical transceiver is depicted in Figure 1.10. It consists of four electrical lanes and optical lanes in each direction, with each lane working at 10G. The lasers are VCSELs with wavelength at 840 nm ∼ 860 nm, and the four optical lanes are carried by four MMFs. The optical transceiver uses 12 fiber multi-parallel optics (MPO) cables and connectors, with four fibers not connected or empty, as shown in Figure 1.10. Although according to IEEE standard, 40G SR4 can only reach 100 m and 150 m over OM3 and OM4 fibers, respectively, the distance of 40G can be significantly improved using some distance-enhancement technologies such as equalization and low-linewidth VCSELs. Up to 300 m and 550 m over OM3 and OM4 fibers can be achieved, respectively, using enhanced 40G SR4 technologies, which cover most of the links in datacenters.