

Communications of **HUAWEI RESEARCH**

March 2023

Issue 4

S+C+L-Band Ultra-high Capacity Optical Transmission p90

Performance Challenges and Key Technologies of
All-Optical Cross-Connect Network Filtering Effect p98

Development and Application of **Optical Fiber Sensing** Technology p132



**LIGHTING THE WAY TOWARD
A FULLY CONNECTED,
FULLY SENSED FUTURE**



Editorial Note

With the large-scale application of ultra-low-loss optical fibers, optical fiber communications has experienced rapid development for more than two decades. It is the best means to provide large-capacity, long-distance information transmission and has become the cornerstone of the information highway, greatly promoting the development of the digital economy. Driven by ultra-large data centers, industry digitalization, and new display technologies, a next-generation optical communications technology system featuring environmental protection, large capacity, low power consumption, and intelligence needs to be researched. In addition, oriented to F5G and future F6G, Huawei will work with industry partners to help drive the optical communications industry and move toward a new era of "Fiber to Everywhere".

Bringing together the wisdom of optical technology researchers, this issue of *Communications of HUAWEI RESEARCH* focuses on long-distance optical transmission, short-distance optical interconnection, and optical access. It presents the latest research progress of core technologies — including optical algorithms, optical amplification, optoelectronic devices, optical systems, and optical cross-connect — and provides an overview and prospect of optical network automatic O&M and optical fiber sensing.

Continuously improving single-fiber capacity is the goal of long-distance optical transmission, while improving the single-wavelength rate and expanding the optical fiber spectrum are key technical paths. *Technology Research on LNOI-based High-Speed Modulators* presents the key technologies and research progress of thin-film lithium niobate modulators with high bandwidth, low driving voltage, and low loss. Furthermore, it demonstrates the significant potential of such modulators in supporting the continuous evolution toward a single-wavelength optical transmission rate higher than 1 Tbit/s. *S+C+L-Band Ultra-high Capacity Optical Transmission* introduces the key challenges and technical breakthroughs of extending the S-band fiber spectrum in theoretical channel modeling, wide-spectrum optical amplification, and optical-layer dynamic performance optimization, and demonstrates that the S+C+L system supports single-fiber 100 Tbit/s optical transmission.

As the foundation of large-scale networking and traffic grooming in optical communication, all-optical cross-connect (OXC) needs to meet the requirements for ultra-large cross-connect capacity, high integration, more degrees, and intelligent O&M management. *Performance Challenges and Key Technologies of All-Optical Cross-Connect Network Filtering Effect* analyzes the source of the filtering effect and the system performance impairment caused by filtering in OXC networks. It also proposes the technical innovation direction for implementing high-bandwidth, low-frequency-offset filtering in OXC networks, facilitating the evolution of OXC networks toward more degrees and higher rates.

High-density integration of optoelectronic devices is a key approach to realize short-distance optical interconnection. *Low-Voltage Silicon Photonic Micro-Ring Modulator Design and Stabilization* presents the research progress in high-performance silicon photonic micro-ring modulators, and discusses how to meet the requirements for high-density, low-power optical interconnection in the future. *Advantages, Progresses, and Applications of Quantum Dot Lasers* introduces the advantages, disadvantage, and application prospect of quantum dot lasers, and elaborates on the latest research progress in directly modulated lasers, high-power O-band lasers, and monolithically integrated lasers grown on silicon. *Antimonide-based High-Sensitivity APDs for 50 Gbit/s Transmission* introduces the high-sensitivity 50 Gbit/s antimonide-based avalanche photodiode (APD) developed by using novel materials.

Home broadband and industry applications pose high requirements on the latency and reliability of optical access networks. The current time division multiplexing mechanism faces major challenges. *New Architecture of x-Dimension Optical Access* proposes a new access architecture from the perspective of multiple access multiplexing and multi-dimension scheduling to meet the requirements of future high-quality networks. Optical fiber sensing can precisely measure multiple physical parameters and is of great significance to industry digitalization. *Development and Application of Optical Fiber Sensing Technology*, a specially invited review paper, provides an overview for point-type, quasi-distributed, and distributed optical fiber sensors based on characteristics of the sensor architecture, providing guidance for the subsequent research on optical fiber sensing technologies.

This issue of *Communications of HUAWEI RESEARCH* also contains many exceptional papers in other research directions, such as Zipper coding technology and optical network performance evaluation and automation. We sincerely hope that these papers provide you with an inspirational read. Should you have any questions, reach out to us anytime for further discussion.



Xinhua Xiao
Huawei Fellow

CONTENTS

Editor-in-Chief:

Heng Liao

Executive Editor:

Xinhua Xiao

Editorial Board:

Heng Liao, Wen Tong, Xinhua Xiao,
Banghong Hu, Huihui Zhou, Feng Bao,
Jeff Xu, Haibo Chen, Pinyan Lu,
Xiaojun Zhang, Ruihua Li, Bo Bai

E-mail: HWResearch@huawei.com

Device

01



Advantages, Progresses, and Applications of Quantum Dot Lasers

Xiangjie Zhao, Xiang Li, Guanlin Lou, Ling Sun, Haihua Qi, Shiyong Zhang

20

Low-Voltage Silicon Photonic Micro-Ring Modulator Design and Stabilization

Dominic J. Goodwill, Patrick Dumais, Jia Jiang, Mohammad M. Rad, Dritan Celso, Jared C. Mikkelsen, Zehua Li, Eric Bernier

38

Technology Research on LNOI-based High-Speed Modulators

Chengcheng Gui, Xiaolu Song, Peijie Zhang, Tao Gui, Shuai Yuan, Keshuang Zheng, Xuanhao Wang, Cheng Zeng, Jinsong Xia

47

Antimonide-based High-Sensitivity APDs for 50 Gbit/s Transmission

Junkai Cao, Wei Xiang, Kai Wang, Wei Wei, Shiyong Zhang, Yan Teng

52

Technical Exploration of Tellurium-based Fiber Amplifiers

Zhipeng Zhao, Zhixu Jia, Weiping Qin, Guanshi Qin, Shiyi Cao, Shengling Wu, Shujie Li, Yehui Liu, Zhiwu Chang, Dejiang Zhang, Xinhua Xiao

System

57



Application of Zipper Code in Optical Fiber Communications

Qinhui Huang, Xiaoling Yang, Huixiao Ma, Kechao Huang

67 • **Coded Modulation with Concatenated Polar-Zipper Codes**
Ali Farsiabi, Hamid Ebrahimzad, Yoones Hashemi, Bashirreza Karimi, Jin Wang, Chuandong Li, Zhuhong Zhang

74 • **Machine Learning for Quality of Transmission Estimation in Optical Networks**
Yvan Pointurier



S+C+L-Band Ultra-high Capacity Optical Transmission
Qiang Guo, Lin Gan, Bo Feng, Shiyi Cao, Xinhua Xiao

98 • **Performance Challenges and Key Technologies of All-Optical Cross-Connect Network Filtering Effect**
Ning Deng, Luo Han, Jingquan Xu, Te'an Lu, Di Xiong, Bowen Zhang, Wei Jia

109 • **New Architecture of x-Dimension Optical Access**
Kuo Zhang, Zhicheng Ye, Borui Li, Liangchuan Li

120 • **ACTN Framework for All-Optical Network Automation and Intent-based Applications**
Henry Yu, Christopher Janz

Sensing



Development and Application of Optical Fiber Sensing Technology
Hao Li, Cunzheng Fan, Xiangpeng Xiao, Liuyang Yang, Liangye Li, Baoqiang Yan, Junfeng Chen, Yuejuan Lv, Haoguang Liu, Yuze Dai, Zhijun Yan, Qizhen Sun



Advantages, Progresses, and Applications of Quantum Dot Lasers

Xiangjie Zhao, Xiang Li, Guanlin Lou, Ling Sun, Haihua Qi, Shiyong Zhang
Gaokun Laboratories, HiSilicon

Abstract

In this paper, we briefly discuss the pros and cons of quantum dot as gain material for laser diodes. By exploiting its advantages, we introduce our progress of quantum dot lasers on three applications: directly modulated laser (DML), high-power O-band laser, and monolithically integrated laser grown on silicon.

Keywords

InAs quantum-dots, lasers, distributed feedback, high power, silicon photonics, integrated optoelectronics

1 Introduction

Not long after the electrically pumped quantum well laser was first demonstrated with much lower threshold than bulk lasers in 1977, researchers considered further improving the laser performance by introducing quantum confinement in more dimensions such as quantum wire (2D confinement) and quantum dot (3D confinement). In 1982, Arakawa and Sakaki theoretically predicted that a laser with three-dimensional confined active media may have temperature-independent threshold current due to the limited carrier thermal spreading of quantum dot discrete state density function as illustrated in Figure 1 [1]. The larger than $2k_B T$ energy gap significantly lowers the probability of the ground state (GS) electrons escaping to higher states, whereas electrons in a bulk or quantum well are free to be redistributed to higher states away from lasing. Since then, lots of theoretical research has been conducted to explore the advantages of quantum dots in the 1980s [2–4]. Asada analyzed the gain characteristics of quantum dot lasers and proposed that the gain of quantum dot can be much higher than quantum well material and bulk material at low bias current, which makes realizing ultra-low threshold quantum dot laser very promising [2]. Furthermore, Arawaka predicted smaller linewidth enhancement factor (LEF) for quantum dot lasers, which may favor reduction of phase noise and lasing linewidth compared to quantum well lasers [5]. In those years, realization of nanometer-scale quantum dots (less than 30 nm in diameter as predicted by Vahala [3]) with high crystalline quality was impossible. Although some of the advantages have been verified experimentally by restricting electrons of conventional lasers in high magnetic fields [1, 6], it was not until the 1990s that researchers figured out a way to synthesize quantum dots directly by epitaxy.

In 1990, Eagesham found in the growth of Ge on Si (100) that the well-known Stranski-Krastanow (SK) growth mode of lattice-mismatched hetero-epitaxy, which was once believed to form an island with misfit dislocation, can actually form dislocation-free islands at the early stage of growth [7]. They named this phenomenon as "coherent SK growth." Coherent SK growth was also observed on InGaAs/GaAs material system [8] and was then used to realize dislocation-free InGaAs quantum dots on GaAs in 1993 [9], and InAs quantum dots on GaAs [10] to achieve O-band emission wavelength. Such quantum dots were called self-assembled quantum dots, implying that the quantum dot can be formed spontaneously on the growth surface

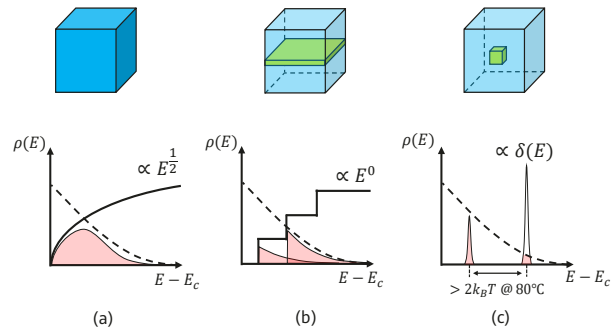


Figure 1 Density of states (DOS) versus electron energy for different free-dimensional structures: (a) Bulk (3D); (b) Quantum well (2D); (c) Quantum dot (0D). The solid lines are DOS, the dashed lines are state occupation probability of electrons, and the shaded areas denote corresponding electron density.

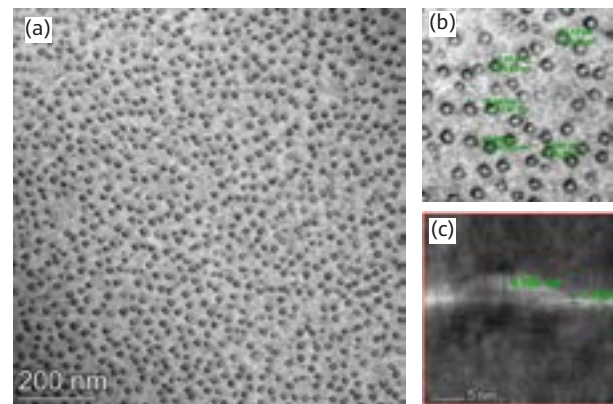


Figure 2 (a) Planar-view TEM of self-assembled quantum dots with lateral size denoted in (b); (c) cross-section TEM image of quantum dots with vertical size denoted

without any pattern-define process. Nowadays, uniformly sized and high-crystal-quality quantum dots can be grown by molecular beam epitaxy (MBE) equipment with atomic layer accuracy as shown in Figure 2, which contains planar-view transmission electron microscopy (TEM) images of InAs quantum dots in GaAs matrix grown by HiSilicon. Those quantum dots have a typical diameter of about 15 nm in the growth plane and 5 nm in the growth direction, and can be adjusted by growth conditions to achieve desired wavelength and DOS for specific application [11].

Various kinds of lasers with quantum dots as active material have been demonstrated since 1994 [12]. The predicted advantages of quantum dot lasers such as high-temperature stability [13], low threshold current [14], and small LEF [15] have also been verified. Unlike the blooming industrial utilization of quantum well lasers, most quantum dot lasers are still in the lab and commercialization cases are fairly rare [16, 17]. While this is mostly due to the early mover advantage of quantum wells, the relatively complicated growth process and some native shortcomings also restrict the implementation of quantum dots.

However, as the bandwidth demands inside data centers are aggressively exploding, the call for integration of high-density and reliable optics with mature silicon electronics industry is pushing quantum dots as a promising laser source and gain solution in future integrated silicon photonics. In section 2, we first discuss the advantages and disadvantages of quantum dots, and then review some of the remarkable achievements from research groups worldwide in the last few decades. In sections 3, 4, and 5, we introduce our progress of quantum dot lasers on reflection insensitive directly modulated laser (DML), high-power O-band lasers, and high performance silicon-based monolithic epitaxial lasers, respectively. While these applications are what we think may be emerging in industrial implementation in a few years to come, the applications of quantum dots are much wider and are under accelerating research efforts. For those who may be interested, several review articles can be referred to [18–23].

2 Quantum Dots as Gain Material: Pros and Cons

2.1 Advantages of Quantum Dots

The first-to-mention change brought by quantum dot to the traditional III-V optoelectronic material system is that it expands the achievable wavelength domain in terms of both gain peak energy and bandwidth. Due to the three-dimensional-island-formation growth mechanism, much larger strain can be relaxed in quantum dot epitaxy than in the layer-by-layer synthesis of quantum wells, without forming any defects [8]. This enables epitaxy of material with a wider lattice for a longer emission wavelength. The direct benefit is making feasible the O-band lasers and amplifiers on 6-inch GaAs substrate by using InAs (6.7% compression strain) as dot material, while conventional O-band lasers are fabricated on 3-inch InP substrate. Considering the massive demands of the O-band laser on short reach intensity-modulation direct-detection (IMDD) scenarios [24, 25], moving to larger substrate may bring a cost advantage to laser diode chip manufacturers.

On the other hand, the lineshape broadening nature of quantum dots is quite different from that of quantum wells in two aspects. First, the self-assemble growth method does not guarantee identical dimension for all the quantum dots. As a result, the GS emission wavelength is slightly different from dot to dot. Each individual dot has its own homogeneous broadening and they are combined together

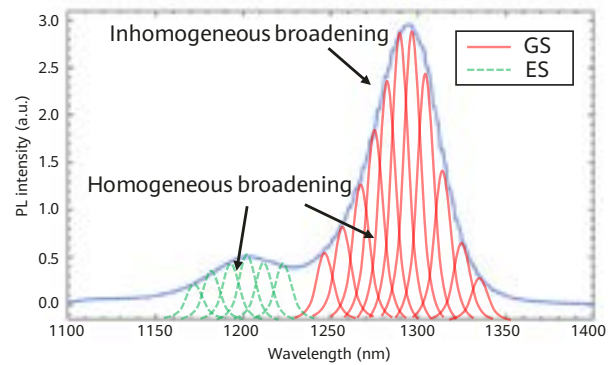


Figure 3 Typical PL spectrum of self-assembled InAs/GaAs quantum dots

to become the Gaussian shaped spectrum (Figure 3) we see in the photoluminescence (PL) test which is called inhomogeneous broadening, whereas the quantum well is dominated by homogeneous broadening. Second and most important, thanks to inhomogeneous broadening, assemblies of separated quantum dots of different GSs can offer gain for light on different wavelength channels with little carrier competition between each other compared to quantum wells, since they are spatially separated. These two features of lineshape broadening are the advantages of quantum dots over quantum wells for wide-band semiconductor optical amplifiers (SOAs) [26, 27] and low-noise mode-locked lasers (MLL) [28, 29].

One of the most well-known advantages of quantum dot lasers is their low threshold and high temperature operation. Many researches have demonstrated abilities of quantum dots, including threshold current as low as 8.4 mA at 100°C in 2002 [30], Fabry-Perot (FP) laser operable up to 220°C and distributed feedback (DFB) laser up to 150°C in 2012 [31]. The low threshold characteristics come from the 3D quantum confinement of the dots. As shown in Figure 2, in each active layer of the laser, quantum dots are randomly distributed islands. Each individual dot possesses two degenerate GS transition pairs of electron and hole. The DOS of each quantum dot is determined by its volume density, which is quite different from a quantum well, whose DOS is estimated from well thickness. For a typical 7 nm thick InAs/GaAs quantum dot layer with area density of $6 \times 10^{10} \text{ cm}^{-2}$, the DOS is estimated to be two orders of magnitude less than the DOS of a single quantum well. This results in less carrier density being required by quantum dots to achieve population inversion, which commonly leads to a lower threshold current. This will also facilitate the quantum dot laser in maintaining lower carrier absorption loss while lasing, making it more energy efficient than its quantum well counterparts [32]. The high temperature

stability comes from another consequence of 3D quantum confinement, the discrete energy states, together with the strong 3D spatial restrict of carriers in dots from interacting with outsiders, bringing about less probability of GS carriers escaping from dots to surrounding barriers. Moreover, for O-band InAs/GaAs quantum dots, the GaAs barrier possesses a band gap as large as InP, making its ability of confining carriers from thermal escape much better than the AlGaInAs quantum well system on InP. These properties make quantum dot an ideal candidate for high-power and energy efficient O-band laser source [33], which we will show in section 4.

LEF, often represented by α or α_H , is defined as the ratio between the derivatives of refractive index with respect to carrier density dn/dN and the differential gain dg/dN , multiplying $-4\pi/\lambda$ [34]. It represents how much the carrier induced gain change affects the refractive index of the active media, which plays a decisive role for the laser to maintain oscillating phase match relation for a certain wavelength. Therefore, a smaller LEF means less phase noise originating from the various carrier fluctuation noises, and high stability under external optical reflection. The carrier induced refractive index change dn/dN is negative for typical quantum well and bulk material, which makes their LEF always positive. For a single quantum dot, the dn/dN can become positive around gain peak, as shown in Figure 4a and Figure 4b, leading to a negative LEF. The physics behind this discriminating feature of quantum dots is atom-like dispersion and Coulomb coupling between discrete and continuum states [35]. More specifically, in a self-assembled InAs/GaAs quantum dot layer with size fluctuation induced inhomogeneous broadening, superposition of in-resonance quantum dots (negative LEF) and off-resonance quantum dots (positive LEF) always results in a smaller LEF than that of quantum wells for each wavelength, at a moderate injection level [36]. Figure 4c shows measured LEF of typical O-band quantum wells and InAs/GaAs quantum dots.

A small LEF can be advantageous to semiconductor lasers from three aspects, i.e., narrow linewidth, isolation-free, and less filamentation. DFB laser with 211 kHz linewidth in O band using InAs/GaAs quantum dot bonded on silicon [37] and 60 kHz in C-band using InAs/InP quantum dots [38] have already been demonstrated. These are much narrower than quantum well DFB lasers which typically have megahertz linewidth. Moreover, tunable laser on silicon on insulator (SOI) platform recently achieved ultra-narrow linewidth of 5.3 kHz by heterogeneously integrating InAs/GaAs quantum dot as gain material [39]. Isolator-

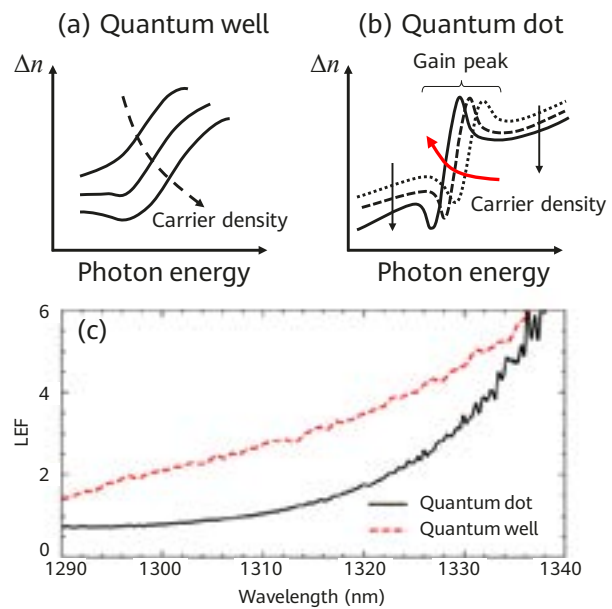


Figure 4 (a) Sketch of carrier induced refractive index change spectra for quantum well or bulk; (b) Sketch of carrier induced refractive index change spectra for quantum dot; (c) Measured LEF of quantum dot and quantum well

free quantum dot DFB lasers may bring cost advantages by removing the need for isolators in traditional optical modules. More importantly, it offers an option for compact on-chip laser source on the SOI photonic integrated circuits (PICs) which often suffer high parasitic reflection due to their high refractive index contrast. Owing to a smaller LEF and a larger damping factor, quantum dot lasers possess higher external reflection tolerance [40] and smaller relative intensity noise (RIN) [41] than quantum well lasers. Various groups have demonstrated their high reflection tolerance quantum dot lasers on GaAs [42–44] and silicon substrates [45, 46]. As shown in the case of an isolator-free 10 Gbit/s DML in section 3, RIN as low as -160 dB/Hz can be achieved. Filamentation is a phenomenon that occurs when the refractive index of waveguide core decreases with increasing injection current due to a positive LEF. The transverse single mode profile can no longer be maintained and a higher transverse mode is stimulated in the laser, causing degradation of laser efficiency and beam quality [47]. It is often encountered in a high-power laser whose waveguide width is typically large for better reliability against catastrophic optical mirror damage (COMD). With a smaller LEF, quantum dot exhibits less of a filamentation problem than quantum well [48], making it a promising candidate for high-power O-band lasers [49, 50].

All points mentioned above are benefits of quantum dots for lasers. When used as gain media for SOAs, quantum dots also exhibit a unique advantage of pattern-effect-free amplification. For conventional quantum well SOAs, after

amplifying the rising edge of the input optical pulse, the depleted carrier reservoir restores inversion in a timescale of about 1 ns, causing a continuous decrease in optical gain through the "1" level for a baud rate of 2.5 Gbit/s or higher. This leads to output signal distortion and raises the bit error rate of a transmission system. However, for a quantum dot SOA, the pattern effect is eliminated thanks to its ultra-fast gain recovery nature (≤ 3 ps) [51–53], which is believed to be because of the fast relaxation of carriers from the first excited state (ES) to the GS due to their high coupling in the 3D confined dot [54]. The much lower density of GS states also ensures that carrier density inversion can be achieved under a moderate injection current. Pattern-effect-free operation [55] of 80 Gbit/s in C-band has already been realized using InAs/InP quantum dots [56], up to 70°C, offering a low-cost compact single chip option for optical regeneration. In addition, even InAs/GaAs quantum dots grown on silicon have demonstrated 60 Gbit/s operation and high power efficiency up to 19.7% for O-band on-chip amplification.

The most impressive advantage of quantum dots is their high tolerance to crystal defects. Lots of groups have published quantum dot laser demos based on heteroepitaxial III-V material on the silicon substrate, in presence of threading dislocation density (TDD) of 10^6 to 10^7 cm⁻² [57–59]. Thousands of hours of high temperature aging of such Si-QD lasers has shown an extrapolated life time of 10 years [60], whereas quantum well lasers under the same TDD only show a life time of several hours with much inferior performance [61]. Such high defect tolerance is based on the material nature of quantum dots [62]. Unlike a quantum well, which is essentially a thin slab of carriers in energy valley where carriers can diffuse freely before being radiatively recombined, each quantum dot is like a "black hole" for carriers confined in the laser active region. As a consequence, the diffusion length is much shorter in quantum dot lasers than in quantum well lasers [63], meaning fewer carriers will be captured by defects to recombine non-radiatively (SRH recombination), which will release energy to the crystal lattice, driving further extension of the existing defects (REDC [64] and REDG [65]). In addition, the low threshold carrier density further reduces the possibility of carriers being captured by defects. The 3D localized wave function in each dot also reduces the possibility of captured carriers escaping to nearby defects. In section 5, we report our recent progress on Si-QD lasers, which show a very low threshold comparable to defect-free quantum dot lasers on a GaAs substrate.

In recent years, Si-QD lasers have become more reliable and are soon expected to meet the requirements for commercial deployment. With tolerance to reflection and defects, quantum dot is now seen as a promising solution for large-scale integration of light sources on silicon chips [66], not only just for the currently available low-cost pluggable silicon photonics modules, but also for future co-packaged optics (CPO) and even inter-chip optical interconnected electronic-photonics integrated circuits (EPICs).

As a brief summary, the advantages of quantum dots as gain material can be categorized into five aspects:

- O-band lasing on GaAs and wide gain spectrum
- Low threshold and high temperature operation
- Small LEF (isolation-free, low noise, and less filamentation)
- Ultra-fast gain recovery
- High tolerance to defects

2.2 Disadvantages of Quantum Dots

Most of the advantages mentioned above originate from the spatial discretization and 3D carrier confinement nature of quantum dots, but these properties also bring two disadvantages that hinder wider application:

- Limited gain
- Limited modulation performance

First, as mentioned above, the GS density of a single layer of self-assembled quantum dots grown by MBE is approximately two orders of magnitude lower than a single quantum well layer. Although this property results in a lower threshold to quantum dot lasers, it limits the maximum available gain of quantum dots, which, according to Fermi's golden rule [67], is directly proportional to DOS. This is why nowadays most quantum dot lasers typically have more than 4 layers to provide enough gain for GS lasing [68]. When the maximum GS gain is not enough to compensate for cavity loss, lasing on ES wavelength will start, which will typically be 50 nm to 100 nm shorter than the GS wavelength. For InAs/GaAs quantum dots, this means that they will no longer lase in the O-band. In addition, due to the poor wave function localization of ESs in a quantum dot, the high temperature carrier injection efficiency of lasing on ES is much lower than on GS [69], leading to an even larger threshold current and lower power efficiency than those of a quantum well. This shortage of gain is more severe in single wavelength lasers, because only parts of the

quantum dot assembly can contribute to the narrow lasing wavelength, meaning a smaller maximum gain. This may put more stringent restrictions on laser design parameters such as waveguide loss, facet reflectivity, and cavity length.

For applications like direct modulation, where a high amount of gain is needed for the short cavity length, the number of quantum dot layers often goes up to 8 or higher. It is important to notice that the number of quantum dot layers cannot continue to increase due to the waveguide design issue. The GaAs spacer between adjacent quantum dot layers is around 30 nm to 40 nm at least, to decouple the strain field influence from the previous dot layer. Therefore, the waveguide core thickness will approach O-band lasing wavelength in GaAs as the layer number goes beyond 8, which may cause difficulties for the epitaxial design to shrink the vertical divergence angle or even to maintain single mode.

Second, the modulation performance is limited for a quantum dot laser. Regardless of C-band quantum dot on InP or O-band quantum dot on GaAs, room-temperature 3 dB bandwidth is generally less than 10 GHz for DFB lasers [70, 71] and less than 12 GHz and 18 GHz for FP lasers on GaAs [72] and InP [73] respectively, both of which are much smaller than those for quantum well counterparts. Large damping is often found on small signal modulation response of these quantum dot lasers, indicating damping limited bandwidth, which is related to the high occupation possibility f_{GS} of electrons in the GS of quantum dots as illustrated by Ishida [74]. The K-factor, which set maximum on 3-dB bandwidth by $f_{3dB,max} = 2\pi\sqrt{2}/K$, can be related to f_{GS} and the relaxation time to GS, τ_0 , as follows:

$$K = 4\pi^2[\tau_p + \tau_0^*] \quad (1)$$

$$= 4\pi^2[\tau_p + \tau_0(1 - f_{GS})^{-1}]$$

Here, τ_p is the photon life time related to all the optical losses in cavity, τ_0^* is the effective electron relaxation time to GS, considering Pauli exclusion principle as $\tau_0^* = \tau_0(1 - f_{GS})^{-1}$. K-factors of 0.14 ns to 0.34 ns are common for quantum well lasers [34, 75], corresponding to bandwidth limit from 63 GHz to 26 GHz, and $\tau_p + \tau_0^*$ about 3.5 ps to 8.6 ps. For quantum dot lasers, f_{GS} is generally larger than that of quantum well lasers because of low DOS and less thermal escaping. The relaxation time of quantum dots (2 ps to 3 ps [52, 74]) τ_0 is also longer than that of quantum wells (0.1 ps to 1 ps [76]). Both factors result in a large damping K-factor (> 0.6 ns) with large f_{GS} playing a major role, corresponding to less than 18 GHz bandwidth. In addition, high speed modulation generally requires a short cavity

length for a smaller RC time constant and photo lifetime τ_p . This often pushes the quantum dot laser to gain saturation regime due to its limited gain, which further lowers the bandwidth through reduced resonance frequency f_R by decreasing the differential gain dg/dN .

In addition, as mentioned above, the spacer thickness between subsequent quantum dot layers is relatively larger than those for quantum wells to decouple the strain impact of previous quantum dot layer. This may lead to higher transport effect in quantum dot lasers, inducing a parasitic roll-off in S21 response, limiting the bandwidth [34]. The higher bandwidth of C-band quantum dot lasers on InP than O-band quantum dot lasers on GaAs may be partly due to reduced transport effect from higher carrier mobility and thinner spacer (20 nm) on the InP platform.

3 Reflection Insensitive 10G DML

In this section, we introduce our work on directly modulated QD-DFB lasers for uncooled and isolator-free 10 Gbit/s applications, such as 10 Gbit/s Ethernet Passive Optical Network (10G-EPON). The fabricated devices achieve an output power of 15 mW at 85°C test environment temperature at 60 mA bias current, and reflection tolerance up to -18 dB with side-mode suppression ratio (SMSR) higher than 50 dB, while maintaining a low RIN of about -160 dB/Hz and an acceptable yield rate.

The device was first grown by MBE with 8 layers of quantum dots. Then a 10 nm InGaAsP was grown by metal organic chemical vapor deposition (MOCVD) as a grating layer, corresponding to a kappa of 21 cm⁻¹ as we measured after fabrication. The grating was patterned by electron beam lithography with a pitch for lasing around 1270 nm and a $\lambda/4$ phase shift was used for improving yield against the random facet phase. A selective wet etch process was used to transfer the grating pattern to InGaAsP against the underlying GaAs. Then the wafer was sent back to MOCVD to grow p-doped GaAs/AlGaAs as the burying upper cladding layers. InGaAsP rather than InGaP was chosen as the grating layer because of its better thermal stability during the burying process. The MOCVD growth temperature was carefully controlled to minimize thermal degradation of quantum dots. After epitaxy, single mode ridge waveguide was formed by dry etching, and contact metal on the surface and rear side were subsequently processed. The chip bar length was cleaved to 650 μ m, then coated with 95% high reflection (HR) and 0.1% AR. After bar testing to

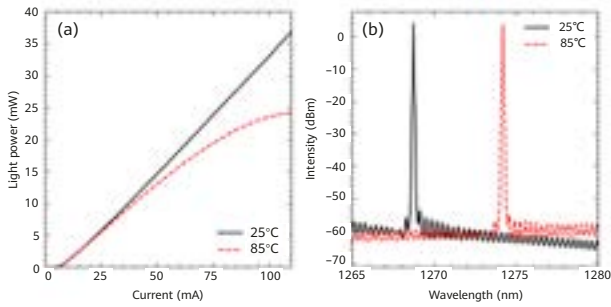


Figure 5 (a) Light-current relation of the 650 μm DFB laser; (b) Lasing spectrum of the 650 μm DFB laser at 60 mA

screen out chips of the bad random facet phase, the remainders were cleaved out and soldered on an aluminium nitride (AlN) substrate for further testing.

The continuous wave (CW) light-current performance and lasing spectrum of the fabricated laser is shown in Figure 5. The threshold current is less than 10 mA from 25°C to 85°C, and is almost unchanged at 25°C and 85°C. At both temperatures, the lasing power is higher than 15 mW with less than 3 mW discrepancy at 60 mA, and the lasing spectrum shows an SMSR larger than 50 dB with a temperature dependent shift of about 0.1 nm/°C. These results show high temperature stability of the QD-DML.

The reflection tolerance of the QD-DML was also tested. Due to limitation of the testing system, the maximum reflection back to the laser facet is -15 dB. It is shown in Figure 6 that an SMSR larger than 50 dB can maintain up to -15 dB reflection at both temperatures. A high yield RIN of around -160 dB/Hz can be realized up to -18 dB reflection at both 25°C and 85°C, as shown in Figure 7. Considering a coupling loss of -4 dB from laser facet to optical fiber, this means the QD-DFB laser can work without an isolator for in-fiber reflection as high as -10 dB.

The small signal 3 dB bandwidth was tested to be around 7.4 GHz and 6.7 GHz at 25°C and 85°C respectively, as shown in Figure 8. From the S21 response, large damping can be clearly recognized as the primary limiting factor of modulation bandwidth as discussed above. Here, we used a shunt capacitance to compensate the high damping. Further improvement can be achieved by decreasing waveguide loss and increasing the maximum gain of quantum dots by high-density dot growth, as reported by K. Takada [70], where a 3 dB-bandwidth of 9.6 GHz at 25°C has been achieved.

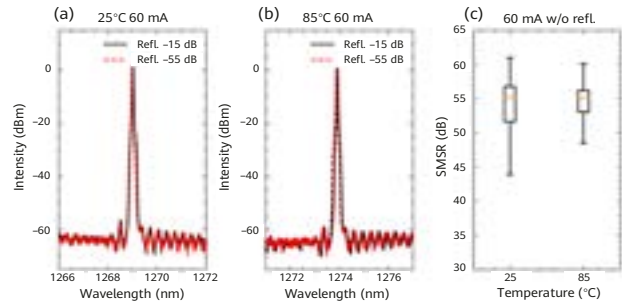


Figure 6 (a) Lasing spectrum with and without -15 dB reflection at 25°C; (b) Lasing spectrum with and without -15 dB reflection at 85°C; (c) SMSR box plot of tested chips

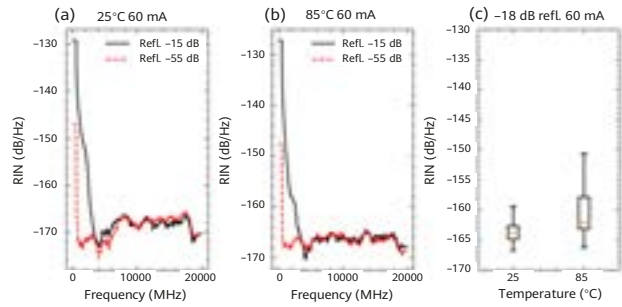


Figure 7 (a) RIN spectrum with and without -15 dB reflection at 25°C; (b) RIN spectrum with and without -15 dB reflection at 85°C; (c) -18 dB reflection RIN box plot of tested chips

4 High-Power O-Band Quantum Dot Laser

As the capacity of electrical switch will grow to 51.2T and higher in the future, bandwidth density required for optical modules will go beyond 800G, to even 1.6T and beyond, making a case for more compact CPO as the solution for balancing power efficiency and bandwidth density [77]. This technology trend is driving the emergence of demands for a high-power and energy efficient O-band laser, where a quantum dot laser on GaAs substrate may exploit its advantage as discussed above. In this section, we will demonstrate our work on a single-mode FP quantum dot laser, which shows wall-plug efficiency (WPE) larger than 26% and output power larger than 100 mW at 85°C and 240 mA.

The device is grown by a single MBE epitaxy on a GaAs substrate, with 4 layers of quantum dots separated by 31 nm GaAs spacer. The confinement factor is designed to be 3%. Single mode ridge waveguide is fabricated by wet etching, resulting in 5.5 μm bottom width and 2.1 μm top

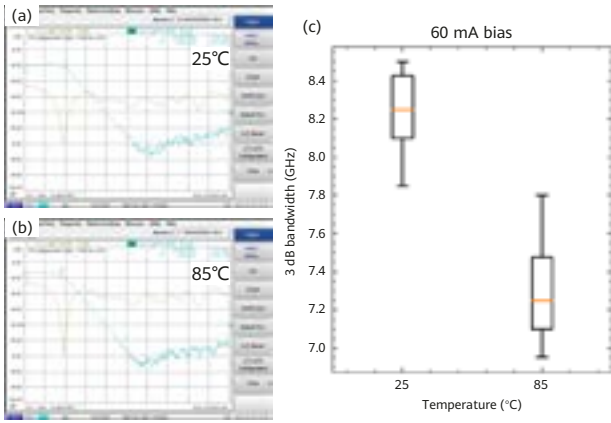


Figure 8 (a) Small signal S21 spectrum at 25°C and 85°C; (b) 3-dB bandwidth box plot of tested chips

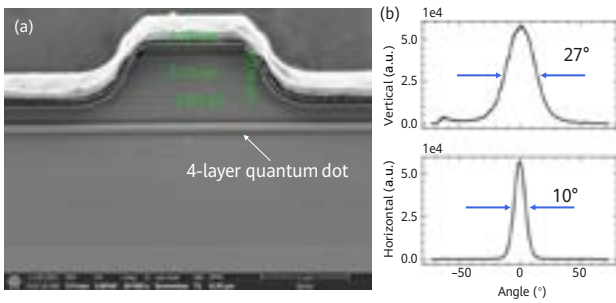


Figure 9 (a) Waveguide structure of the quantum dot laser wet etched to show the epitaxial layer structures; (b) Laser beam far field on vertical and horizontal directions

width, as shown in Figure 9a. Electron beam evaporation followed by plating is used to form the p-side metal for better heat dissipation. After fabrication, the wafer is cleaved into bars of different lengths, then coated with HR (95%) on one facet and different ARs on the other facet to investigate gain capacity of quantum dots for different cavity losses. The divergence angles are 27° and 10° for the vertical and horizontal directions, as shown in Figure 9b. Because the maximum gain of quantum dots is relatively small, the facet out-coupling efficiency and loss need to be carefully adjusted for high power efficiency.

All the devices were tested under the CW condition. The median bar threshold current density versus facet loss $\alpha_m = \frac{1}{L} \ln\left(\frac{1}{r_1 r_2}\right)$ is plotted in Figure 10a. The test temperature is 85°C. It can be seen that, below a cavity loss of 7.8 cm⁻¹, the threshold current density slowly increases as cavity loss increases, indicating a high differential gain. However, after the loss exceeds 8 cm⁻¹, the threshold current density jumps to a value higher than 1000 A/cm², indicating an abrupt degradation of differential gain. This is the gain limitation discussed above, and is distinctively unlike the quantum well whose differential gain decreases relatively slowly with increasing loss. It should be noticed that, once the gain

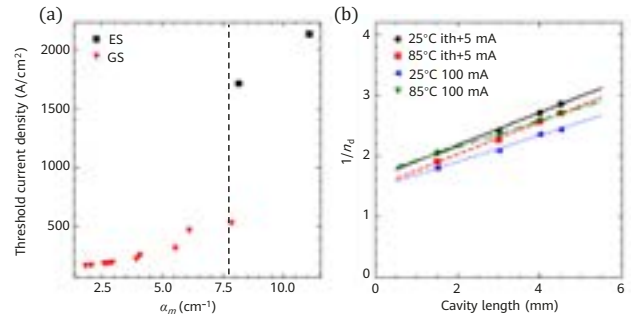


Figure 10 (a) Threshold current density as a function of nominal facet loss; (b) Fitted $1/\eta_d$ versus cavity length

limitation is exceeded, the promised advantages of quantum dots, including high efficiency, small LEF, and tolerance to defects, will degrade rapidly.

Current injection efficiency η_i and internal loss α_i are fitted based on differential efficiency $1/\eta_d$ of various cavity lengths, all coated by 20% AR. Samples used for fitting are all below gain limitation. The fitted line is plotted in Figure 10b. The internal loss is estimated to be around 1.6 cm⁻¹. All fitted lines have R-squared larger than 99%. It is shown that a high injection efficiency of 73.2% can be achieved, despite the dots only covering 12% of the layer area. Interestingly, η_i at current of 100 mA is larger than that just above threshold. This may be due to the enhanced carrier-carrier scattering, which decreases the carrier relaxation life time from higher states to the GS of quantum dots. This indicates that the carrier density above threshold in the quantum dot is not well clamped like in the quantum well, due to a relatively large carrier relaxation time to quantum dot GS [78]. The best WPE achieved in those samples is 1500 μm cavity length with 15% AR coating. The output power and WPE versus current are plotted in Figure 11. Samples with 15% AR coating show maximum WPE of 34.3% and 26.4%, at 25°C and 85°C, respectively. The output power at 400 mA reaches 212 mW and 153 mW at 25°C and 85°C, respectively. The optical power kink in sample with 20% AR coating at 25°C is a consequence of higher transverse mode due to the large lower waveguide width. Above 300 mA at 85°C, the start of power rollover can be seen in both samples. Compared to samples with 20% AR coating, the 15% AR coating samples show higher threshold current and more apparent tendency of rollover, indicating that the quantum dots are approaching the gain limited regime with increasing carrier density and temperature of the active region. For future implementation of quantum dots in DFB lasers, more quantum dot layers are recommended to obtain a higher margin of gain at high temperature, because only parts of the dots can provide gain for a single wavelength.

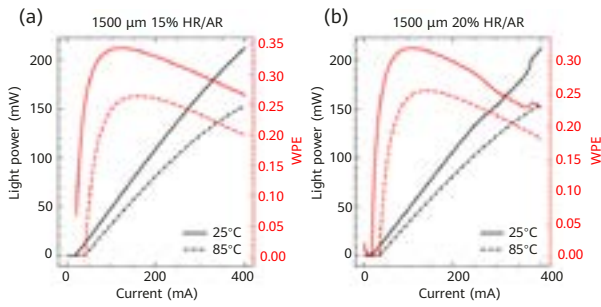


Figure 11 (a) Output power and WPE versus current of 1500 μm cavity length with 15% AR coating; (b) Output power and WPE versus current of 1500 μm cavity length with 20% AR coating

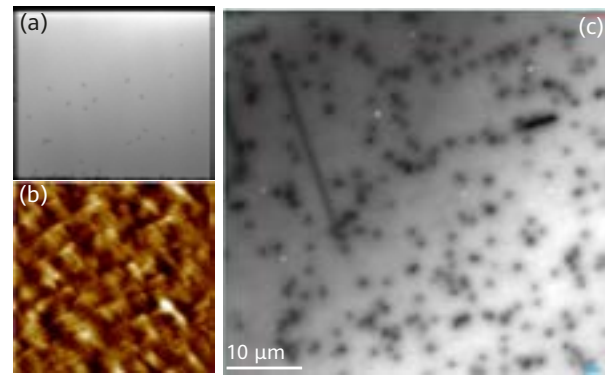


Figure 12 (a) Plan-view TEM image of the GaAs on silicon sample with image size of 10 μm x 10 μm; (b) AFM image of the GaAs on silicon sample with image size of 10 μm x 10 μm; (c) CL image of 50 μm x 50 μm after superlattice

5 Monolithic Quantum Dot Laser Grown on Silicon

Silicon photonics is growing as a powerful competitor to traditional optical devices based on the III-V material system, due to its compactness, compatibility to complementary metal-oxide-semiconductor (CMOS) massive production, and promising wafer level integration with electronic silicon components [79]. Nevertheless, the indirect band gap nature of silicon means that it is not an efficient material for light generation. Various efforts have been made to integrate light sources on silicon in the last two decades [80–82], and the most successful route up to now is the bonding of III-V quantum well to SOI waveguide [83], which has already been launched to the market [84]. However, separated III-V wafer and pick-align-put process are still needed in the bonding procedure, limiting both yield and throughput [85]. Coming closer to a fully integrated wafer process, directly growing III-V laser material on silicon has been regarded as the ultimate solution [66]. To reduce dislocation density caused by large thermal, lattice, and polarity mismatches between the silicon substrate and III-V layer, several methods have been commonly adopted, including Ge or GaP intermediate buffer layers [86, 87], large off-cut angle, or patterned Si (001) wafers [88–90]. However, off-cut substrates are incompatible with the standard CMOS fabrication process, which demands nominal on-axis Si (001) substrates. Intermediate buffer layers and patterned substrates will increase process complexity, especially for monolithic integration on SOI substrates, and thus reduce the feasibility of manufacturing. In this section, we will report our work on InAs/GaAs self-assembled quantum dot lasers monolithically grown on CMOS-compatible, nominal on-axis Si (001) substrate without any intermediate buffer layers or patterns.

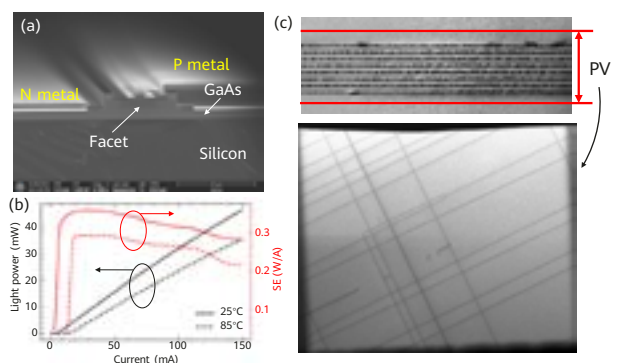


Figure 13 (a) SEM image of the cleaved facet of Si-QD laser; (b) Output power and slope efficiency as a function of current; (c) Cross section and planar view TEM image of the laser's quantum dot epitaxial layers

The InAs/GaAs quantum dot epitaxy is grown on a 3-inch silicon wafer. The first obstacle in the way of growing GaAs on silicon is the anti-phase boundary problem, due to the polar/non-polar epitaxy interface [91]. We resolved this problem by high temperature annealing of the substrate in H₂ atmosphere by MOCVD to reconstruct the double atomic step on the silicon surface, preventing the formation of the anti-phase boundary. In the subsequent GaAs epitaxy, a three-step growth method was employed. First, Tertiarybutylarsine (TBAs) was used instead of AsH₃ for low temperature nucleation layer to relax most of the strain due to lattice mismatch between silicon and GaAs. Then, we switched to AsH₃ to deposit another 100 nm GaAs layer at a mid-temperature, followed by a 1000 nm GaAs layer grown at a high temperature to complete anti-phase domain-free buffer growth. The growth condition was carefully adjusted to prevent stacking faults. In addition, several thermal cycle annealing processes were also inserted during high temperature GaAs layer growth to improve the crystal quality, judged by PL, and to decrease the TDD, judged by plan-view-TEM and cathodoluminescence (CL). The resulting GaAs on Silicon template has a GaAs thickness

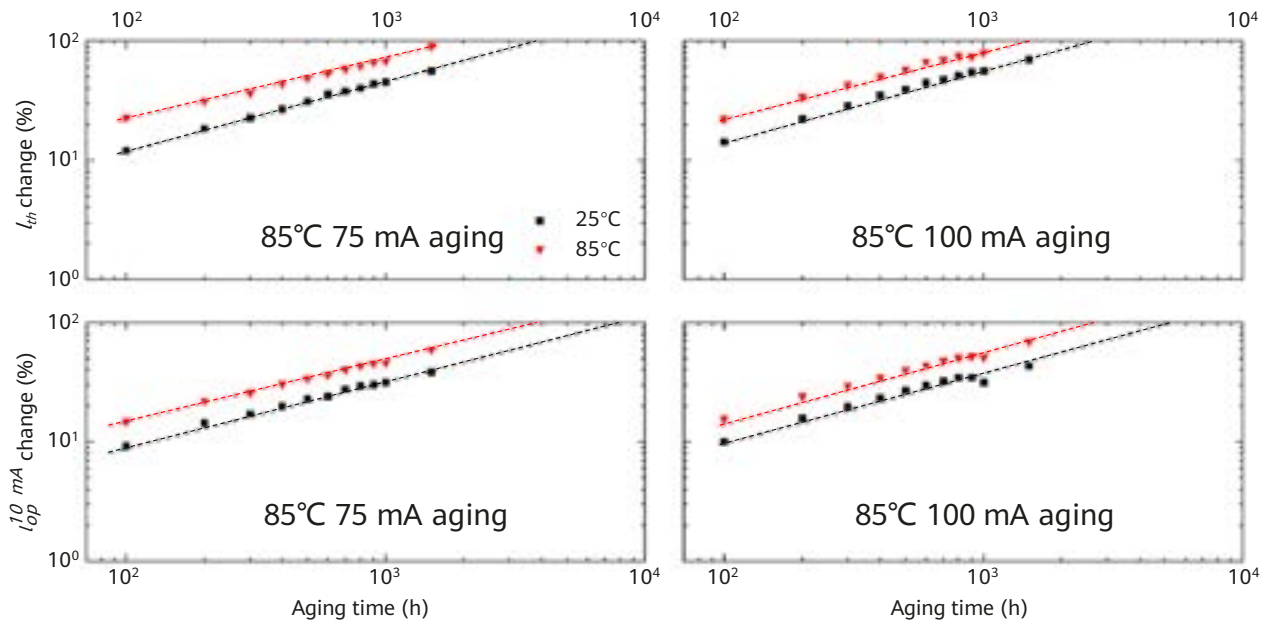


Figure 14 Median percentage degradation of threshold current and bias for 10 mW output power under aging condition of 85°C 75 mA and 100 mA. The dashed lines are guided for the eye.

of 1.2 μm and TDD of $2\text{e}7\text{ cm}^{-2}$, as shown in Figure 12. It also shows the root mean square roughness is 0.623 nm measured by atomic force microscopy (AFM). Further, TDD filtering was performed by two sets of superlattice, which end up with a TDD of $7\text{e}6\text{ cm}^{-2}$ as confirmed by CL in Figure 12c.

The GaAs on silicon sample was then sent into MBE to grow a quantum dot laser structure. Eight layers of quantum dots were utilized with the resulting total thickness of III-V layers on silicon being 7 μm . The as-grown material was then processed into single mode waveguide FP laser with a ridge width of 1.8 μm . Co-planar N-contact was used instead of back-metal. The wafer was thinned and polished to a thickness of 120 μm and cleaved to bars of 1000 μm cavity length. One facet of the bars were coated with HR (95%) film, whereas the other facet was left as cleaved (AC). Figure 13a shows the facet image of the laser bars. Note that although the facet of silicon part is rough, the III-V facet remains as a perfectly cleaved crystal plane.

The device light-current performance was measured under the CW condition. One of the results is shown in Figure 13b. Very low threshold current of 5.4 mA and 16.7 mA are realized at 25°C and 85°C, respectively, which is comparable to native GaAs quantum dot devices. The device also achieves maximum slope efficiency of 35.8% and 29.3% and output power at 100 mA of 32 mW and 24 mW at each temperature. The injection efficiency and internal loss extracted from median $1/\eta_d$ near the threshold of 650 μm , 750 μm , and 1000 μm lasers are 45.2% and 4.39 cm^{-1} ,

respectively. The relatively low injection efficiency is partly due to the misfit dislocation network in the quantum dot epitaxial layers, as shown by the cross section and planar view TEM images of the quantum dot region in Figure 13c. These misfit dislocation networks are formed primarily during the cooling after epitaxy growth, because of thermal expansion coefficient mismatch between III-V and silicon [63]. They can be moved out from the quantum dot layer by introducing trapping layers located in AlGaAs cladding just above and below the quantum dot region [65, 92]. We have also fabricated lasers with such trapping layers, which show an improvement in injection efficiency of up to 63.7%.

As for the reliability performance, chips whose 85°C threshold current is less than 20 mA were diced from bars and mounted on AlN carriers for aging. Two aging conditions were selected, 85°C 75 mA and 100 mA, each with 17 chips. Every 100 hours, the chip were taken out of the aging rack and tested for light-current relations at 25°C and 85°C. The results of median percentage degradation are shown in Figure 14 for threshold current and bias for 10 mW output power. No sudden failure was observed during the 1500 hours of aging. All the chips suffered a large degradation in the first 100 hours. This may be primarily attributed to the recombination enhanced dislocation climb (REDC) [64] of the misfit dislocation network in quantum dot layers. As the aging proceeds, the degradation tends to slow down, especially in the case of 85°C 100 mA. Better reliability can be realized by QD-on-Si lasers with the trapping layers

mentioned earlier [60]. Our results for such kind of lasers will be reported once the reliability tests are finished.

6 Summary and Future Prospects

Since it was first proposed in 1982, followed by epitaxial realization in the early 1990s, quantum dots as gain material for lasers and optical amplifiers have undergone intensive researches, both theoretically [93–95] and experimentally [96, 97]. Their advantages, including high temperature operation, low noise, and tolerance to reflection and crystal defects, are becoming well known in the industry and spawning exciting innovative applications in the academia [78]. But frankly speaking, implementation of quantum dot technologies is still new for the industrial community even until now, because no exclusive application scenario has been found in the last two decades. The main reason for this may be that in the era of pluggable optics, those advantages are not attractive enough to lever-age the industrial resource from their mature quantum well counterparts, because quantum well devices can still meet the requirements. However, as the density and scale of optical components continue to grow and they are more closely integrated with electronic silicon processors, the power budget for lasers is squeezed. Also, the reflection and environment temperature are much harsher for on-chip light sources than in pluggable modules. The demands for high efficiency as an external laser or the demands for isolator-free and high reliability as an on-chip light source point to the advantages of quantum dots. Given the maturity of the design and process of high-power lasers on the GaAs material system and the low cost of 6-inch GaAs wafers, the emergence of commercial high-power and energy efficient QD-DFB lasers is expected to be rapid. After fulfillment of its power efficiency and reflection tolerance at high output power and high temperature, hybrid [82] and heterogeneous [83] integration of quantum dots for on-chip light sources may have more chances for industrial development. There is already a demonstration of silicon photonic micro-transceiver with hybrid integrated quantum dot lasers for operation at ambient temperature above 105°C [98].

The epitaxial integration of quantum dot with silicon photonics, once realized, will undoubtedly out-compete any other integration scheme in terms of material costs and scalability. Two technical obstacles are still stand in the way. One is the reliability. Various groups, including ours, have rapidly progressed from the reliability perspective. We

expect to resolve this problem in the coming years. The other obstacle is the coupling scheme between the grown quantum dot lasers and SOI waveguide. There were several proposals in literature [20, 83, 99], but this has been rarely realized experimentally. One of the major difficulties, in our perspective, is that, to achieve high coupling efficiency and compatibility with CMOS processes, selective regrowth of quantum dot lasers on epitaxial windows of SOI wafer is necessary, and the growth profile at those epitaxial window edges needs to be free of the facet effect [100] for competitive efficiency of either evanescent or butt coupling. However, none of these can be easily accomplished by MBE, the conventional epitaxial equipment for quantum dots. MOCVD can possibly solve those difficulties [101] with the ability for selective epitaxy, growth morphology control [101], and quantum dot growth [102]. But all these technologies are in their early stages, and more research effort is required in these areas to prove the feasibility. Another difficulty is the thermal compatibility of the high double-step annealing temperature, typically higher than 950°C, with silicon photonic doping profiles, silicides, and germanium. Germanium photodiode may be replaced by InGaAs on Si to partly overcome this problem.

In summary, we briefly reviewed both the advantages and disadvantages of quantum dots as gain material. Our work with quantum dots on 10G DML, high-power O-band laser, and silicon-based epitaxial lasers were also reported. Quantum dots are expected to best exert their advantages in terms of high efficiency, high temperature operation, high tolerance to reflection, and high reliability as a light source for large-scale silicon photonics systems.

References

- [1] Y. Arakawa and H. Sakaki, "Multidimensional quantum well laser and temperature dependence of its threshold current," *Applied Physics Letters*, vol. 40, no. 11, pp. 939–941, 1982. [Online]. Available: <https://doi.org/10.1063/1.92959>
- [2] M. Asada, Y. Miyamoto, and Y. Suematsu, "Gain and the threshold of three-dimensional quantum-box lasers," *IEEE Journal of Quantum Electronics*, vol. 22, no. 9, pp. 1915–1921, Sep. 1986.
- [3] K. Vahala, "Quantum box fabrication tolerance and size limits in semiconductors and their effect on optical gain," *IEEE Journal of Quantum Electronics*, vol. 24, no. 3, pp. 523–530, March 1988.
- [4] Y. Miyamoto, Y. Miyake, M. Asada, and Y. Suematsu, "Threshold current density of GaInAsP/InP quantum-box lasers," *IEEE Journal of Quantum Electronics*, vol. 25, no. 9, pp. 2001–2006, Sep. 1989.
- [5] Y. Arakawa, K. Vahala, and A. Yariv, "Dynamic and spectral properties of semiconductor lasers with quantum-well and quantum-wire effects," *Surface Science*, vol. 174, no. 1, pp. 155–162, 1986.
- [6] K. Vahala, Y. Arakawa, and A. Yariv, "Reduction of the field spectrum linewidth of a multiple quantum well laser in a high magnetic field—spectral properties of quantum dot lasers," *Applied Physics Letters*, vol. 50, no. 7, pp. 365–367, 1987. [Online]. Available: <https://doi.org/10.1063/1.98200>
- [7] Eaglesham and Cerullo, "Dislocation-free Stranski-Krastanow growth of Ge on Si(100)," *Physical review letters*, vol. 64, no. 16, pp. 1943–1946, 1990.
- [8] C. W. Snyder, B. G. Orr, D. Kessler, and L. M. Sander, "Effect of strain on surface morphology in highly strained InGaAs films," *Phys. Rev. Lett.*, vol. 66, pp. 3032–3035, Jun 1991. [Online]. Available: <https://link.aps.org/doi/10.1103/PhysRevLett.66.3032>
- [9] D. Leonard, M. Krishnamurthy, C. M. Reaves, S. P. Denbaars, and P. M. Petroff, "Direct formation of quantum-sized dots from uniform coherent islands of InGaAs on GaAs surfaces," *Applied Physics Letters*, vol. 63, no. 23, pp. 3203–3205, 1993. [Online]. Available: <https://doi.org/10.1063/1.110199>
- [10] K. Nishi, H. Saito, S. Sugou, and J.-S. Lee, "A narrow photoluminescence linewidth of 21 meV at 1.35 μm from strain-reduced InAs quantum dots covered by $\text{In}_{0.2}\text{Ga}_{0.8}\text{As}$ grown on GaAs substrates," *Applied Physics Letters*, vol. 74, no. 8, pp. 1111–1113, 1999. [Online]. Available: <https://doi.org/10.1063/1.123459>
- [11] Y. Nakata, K. Mukai, M. Sugawara, K. Ohtsubo, H. Ishikawa, and N. Yokoyama, "Molecular beam epitaxial growth of InAs self-assembled quantum dots with light-emission at 1.3 μm ," *Journal of Crystal Growth*, vol. 208, no. 1, pp. 93–99, 2000.
- [12] N. Kirstaedter, N. Ledentsov, M. Grundmann, D. Bimberg, V. Ustinov, S. Ruvimov, M. Maximov, P. Kop'ev, Z. Alferov, U. Richter, P. Werner, U. Gösele, and J. Heydenreich, "Low threshold, large T0 injection laser emission from InGaAs quantum dots," *Electronics Letters*, vol. 30, pp. 1416–1417(1), August 1994.
- [13] S. Fathpour, Z. Mi, P. Bhattacharya, A. R. Kovsh, S. S. Mikhlin, I. L. Krestnikov, A. V. Kozhukhov, and N. N. Ledentsov, "The role of Auger recombination in the temperature-dependent output characteristics ($T_0=\infty$) of p-doped 1.3 μm quantum dot lasers," *Applied Physics Letters*, vol. 85, no. 22, pp. 5164–5166, 2004. [Online]. Available: <https://doi.org/10.1063/1.1829158>
- [14] K. Mukai, Y. Nakata, K. Ohtsubo, M. Sugawara, N. Yokoyama, and H. Ishikawa, "Low threshold (8 mA) 1.3 μm CW lasing of InGaAs/InAs quantum dots at room temperature," in *Conference on Lasers and Electro-Optics*, ser. OSA Technical Digest, K. W. K. J. Chang-Hasnain, C. and K. Vahala, Eds. Baltimore, Maryland: Optical Society of America, 1999, p. CPD18. [Online]. Available: <http://www.osapublishing.org/abstract.cfm?URI=CLEO-1999-CPD18>
- [15] Jungho Kim and Shun Lien Chuang, "Theoretical and experimental study of optical gain, refractive index change, and linewidth enhancement factor of p-doped quantum-dot lasers," *IEEE Journal of Quantum Electronics*, vol. 42, no. 9, pp. 942–952, Sep. 2006.
- [16] "Innolume GmbH," <https://www.innolume.com/products/>, 2021.

- [17] "QD Laser Co., Inc." <https://qdlaser.com/en/products/list.html>, 2021.
- [18] Y. Wan, J. Norman, S. Liu, A. Liu, and J. E. Bowers, "Quantum dot lasers and amplifiers on silicon: Recent advances and future developments," *IEEE Nanotechnology Magazine*, vol. 15, no. 2, pp. 8–22, April 2021.
- [19] S. Bauer, V. Sichkovskiy, O. Eyal, T. Septon, A. Becker, I. Khanonkin, G. Eisenstein, and J. P. Reithmaier, "1.5 μm indium phosphide-based quantum dot lasers and optical amplifiers: The impact of atom-like optical gain material for optoelectronics devices," *IEEE Nanotechnology Magazine*, vol. 15, no. 2, pp. 23–36, April 2021.
- [20] J. C. Norman, D. Jung, Y. Wan, and J. E. Bowers, "Perspective: The future of quantum dot photonic integrated circuits," *APL Photonics*, vol. 3, no. 3, p. 030901, 2018. [Online]. Available: <https://doi.org/10.1063/1.5021345>
- [21] J. C. Norman, R. P. Mirin, and J. E. Bowers, "Quantum dot lasers—History and future prospects," *Journal of Vacuum Science & Technology A*, vol. 39, no. 2, p. 020802, 2021. [Online]. Available: <https://doi.org/10.1116/6.0000768>
- [22] S. Pan, V. Cao, M. Liao, Y. Lu, Z. Liu, M. Tang, S. Chen, A. Seeds, and H. Liu, "Recent progress in epitaxial growth of III–V quantum-dot lasers on silicon substrate," *Journal of Semiconductors*, vol. 40, no. 10, p. 101302, Oct. 2019. [Online]. Available: <https://doi.org/10.1088/1674-4926/40/10/101302>
- [23] Y. Arakawa, "Progress in quantum dots for advanced photonics," in *2017 22nd Microoptics Conference (MOC)*, Nov 2017, pp. 14–15.
- [24] K. Takada, Y. Tanaka, T. Matsumoto, M. Yamaguchi, T. Kageyama, K. Nishi, Y. Nakata, T. Yamamoto, M. Sugawara, and Y. Arakawa, "High-output-power 10.3 Gb/s Operation of 1.27 μm Quantum-dot DFB Lasers for 10G-EPON," in *CLEO:2011 - Laser Applications to Photonic Applications*. Optical Society of America, 2011, p. CFD5.
- [25] K. Kurata, Y. Hagihara, M. Kurihara, K. Yashiki, K. Kinoshita, K. Shiba, M. Kuwata, T. Muto, S. Kobayashi, N. Baba, K. Fujii, and R. Pitwon, "Short reach, high temperature operation and high reliability silicon photonic micro-transceivers for embedded and co-packaged system integration," in *Optical Interconnects XXI*, H. Schröder and R. T. Chen, Eds., vol. 11692, International Society for Optics and Photonics. SPIE, 2021, pp. 1–16. [Online]. Available: <https://doi.org/10.1117/12.2576670>
- [26] T. Akiyama, M. Ekawa, M. Sugawara, K. Kawaguchi, Hisao Sudo, A. Kuramata, H. Ebe, and Y. Arakawa, "An ultrawide-band semiconductor optical amplifier having an extremely high penalty-free output power of 23 dBm achieved with quantum dots," *IEEE Photonics Technology Letters*, vol. 17, no. 8, pp. 1614–1616, Aug 2005.
- [27] O. Eyal, A. Willinger, S. Banyoudeh, F. Schanbel, V. Sichkovskiy, V. Mikhelashvili, J. P. Reithmaier, and G. Eisenstein, "Static and dynamic characteristics of an InAs/InP quantum-dot optical amplifier operating at high temperatures," *Opt. Express*, vol. 25, no. 22, pp. 27 262–27 269, Oct 2017. [Online]. Available: <http://www.osapublishing.org/oe/abstract.cfm?URI=oe-25-22-27262>
- [28] G. Carpintero, M. G. Thompson, K. Yvind, R. V. Penty, and I. H. White, "Comparison of the noise performance of 10 GHz repetition rate quantum-dot and quantum well monolithic mode-locked semiconductor lasers," *IET Optoelectronics*, vol. 5, no. 5, pp. 195–201, Oct. 2011.
- [29] M. AL-QADI, M. O'Sullivan, C. Xie, and R. Hui, "Phase noise spectral properties across individual comb lines in quantum-dot mode-locked lasers," in *Optical Fiber Communication Conference (OFC) 2020*. Optical Society of America, 2020, p. M1H.2. [Online]. Available: <http://www.osapublishing.org/abstract.cfm?URI=OFC-2020-M1H.2>
- [30] O. Shchekin and D. Deppe, "Low-threshold high-T0 1.3 μm InAs quantum-dot lasers due to p-type modulation doping of the active region," *IEEE Photonics Technology Letters*, vol. 14, no. 9, pp. 1231–1233, Sep. 2002.
- [31] T. Kageyama, K. Takada, K. Nishi, M. Yamaguchi, R. Mochida, Y. Maeda, H. Kondo, K. Takemasa, Y. Tanaka, T. Yamamoto, M. Sugawara, and Y. Arakawa, "Long-wavelength quantum dot FP and DFB lasers for high temperature applications," in *Novel In-Plane Semiconductor Lasers XI*, A. A. Belyanin and P. M.

- Smowton, Eds., vol. 8277, International Society for Optics and Photonics. SPIE, 2012, pp. 73–80. [Online]. Available: <https://doi.org/10.1117/12.905873>
- [32] L. V. Asryan, S. Luryi, and R. A. Suris, "Internal efficiency of semiconductor lasers with a quantum-confined active region," *IEEE Journal of Quantum Electronics*, vol. 39, no. 3, pp. 404–418, March 2003.
- [33] S. S. Mikhlin, A. R. Kovsh, I. L. Krestnikov, A. V. Kozhukhov, D. A. Livshits, N. N. Ledentsov, Y. M. Shernyakov, I. I. Novikov, M. V. Maximov, V. M. Ustinov, and Z. I. Alferov, "High power temperature-insensitive 1.3 μm InAs/InGaAs/GaAs quantum dot lasers," *Semiconductor Science and Technology*, vol. 20, no. 5, pp. 340–342, mar 2005. [Online]. Available: <https://doi.org/10.1088/0268-1242/20/5/002>
- [34] S. W. C. Larry A. Coldren and M. L. Mašanović, *Dynamic Effects*. John Wiley Sons, Ltd, 2012, ch. Five, pp. 247–333. [Online]. Available: <https://doi.org/10.1002/9781118148167.ch5>
- [35] H. C. Schneider, W. W. Chow, and S. W. Koch, "Anomalous carrier-induced dispersion in quantum-dot active media," *Phys. Rev. B*, vol. 66, p. 041310, Jul 2002. [Online]. Available: <https://link.aps.org/doi/10.1103/PhysRevB.66.041310>
- [36] Z. Zhang, D. Jung, J. C. Norman, W. W. Chow, and J. E. Bowers, "Linewidth enhancement factor in InAs/GaAs quantum dot lasers and its implication in isolator-free and narrow linewidth applications," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 25, no. 6, pp. 1–9, 2019.
- [37] D. Liang, S. Srinivasan, A. Descos, C. Zhang, G. Kurczveil, Z. Huang, and R. Beausoleil, "High-performance quantum-dot distributed feedback laser on silicon for high-speed modulations," *Optica*, vol. 8, no. 5, pp. 591–593, May 2021.
- [38] T. Septon, A. Becker, S. Gosh, G. Shtendel, V. Sichkovskiy, F. Schnabel, A. Sengül, M. Bjelica, B. Witzigmann, J. P. Reithmaier, and G. Eisenstein, "Large linewidth reduction in semiconductor lasers based on atom-like gain material," *Optica*, vol. 6, no. 8, pp. 1071–1077, Aug 2019.
- [39] A. Malik, J. Guo, M. A. Tran, G. Kurczveil, D. Liang, and J. E. Bowers, "Widely tunable, heterogeneously integrated quantum-dot O-band lasers on silicon," *Photon. Res.*, vol. 8, no. 10, pp. 1551–1557, Oct 2020. [Online]. Available: <http://www.osapublishing.org/prj/abstract.cfm?URI=prj-8-10-1551>
- [40] G. Huyet, D. O'Brien, S. Hegarty, J. Mcinerney, A. Uskov, D. Bimberg, C. Ribbat, V. Ustinov, A. Zhukov, S. Mikhlin, A. Kovsh, J. White, K. Hinzer, and A. Springthorpe, "Quantum dot semiconductor lasers with optical feedback," *physica status solidi (a)*, vol. 201, pp. 345–352, 01 2004.
- [41] A. Capua, L. Rozenfeld, V. Mikhelashvili, G. Eisenstein, M. Kuntz, M. Laemmlin, and D. Bimberg, "Direct correlation between a highly damped modulation response and ultra low relative intensity noise in an InAs/GaAs quantum dot laser," *Opt. Express*, vol. 15, no. 9, pp. 5388–5393, Apr 2007. [Online]. Available: <http://www.osapublishing.org/oe/abstract.cfm?URI=oe-15-9-5388>
- [42] B. Dong, J. Duan, H. Huang, J. C. Norman, K. Nishi, K. Takemasa, M. Sugawara, J. E. Bowers, and F. Grillot, "Dynamic performance and reflection sensitivity of quantum dot distributed feedback lasers with large optical mismatch," *Photon. Res.*, vol. 9, no. 8, pp. 1550–1558, Aug 2021. [Online]. Available: <http://www.osapublishing.org/prj/abstract.cfm?URI=prj-9-8-1550>
- [43] O. Carroll, I. O'Driscoll, S. P. Hegarty, G. Huyet, J. Houlihan, E. A. Viktorov, and P. Mandel, "Feedback induced instabilities in a quantum dot semiconductor laser," *Opt. Express*, vol. 14, no. 22, pp. 10 831–10 837, Oct 2006. [Online]. Available: <http://www.osapublishing.org/oe/abstract.cfm?URI=oe-14-22-10831>
- [44] N. Hatori, K. Mizutani, S.-H. Jeong, Y. Tanaka, and K. Kurata, "High reflection tolerance of quantum dot distributed feedback lasers for silicon photonics transmitters," in *2017 IEEE 14th International Conference on Group IV Photonics (GFP)*, Aug 2017, pp. 105–106.
- [45] H. Huang, J. Duan, B. Dong, J. Norman, D. Jung, J. E. Bowers, and F. Grillot, "Epitaxial quantum dot lasers on silicon with high thermal stability and strong resistance to optical feedback," *APL Photonics*,

- vol. 5, no. 1, p. 016103, 2020. [Online]. Available: <https://doi.org/10.1063/1.5120029>
- [46] J. Duan, H. Huang, B. Dong, D. Jung, J. C. Norman, J. E. Bowers, and F. Grillot, "1.3- μm reflection insensitive InAs/GaAs quantum dot lasers directly grown on silicon," *IEEE Photonics Technology Letters*, vol. 31, no. 5, pp. 345–348, March 2019.
- [47] J. Marciante and G. Agrawal, "Nonlinear mechanisms of filamentation in broad-area semiconductor lasers," *IEEE Journal of Quantum Electronics*, vol. 32, no. 4, pp. 590–596, April 1996.
- [48] P. M. Smowton, E. J. Pearce, H. C. Schneider, W. W. Chow, and M. Hopkinson, "Filamentation and linewidth enhancement factor in InGaAs quantum dot lasers," *Applied Physics Letters*, vol. 81, no. 17, pp. 3251–3253, 2002. [Online]. Available: <https://doi.org/10.1063/1.1516236>
- [49] E. Gehrig, O. Hess, C. Ribbat, R. L. Sellin, and D. Bimberg, "Dynamic filamentation and beam quality of quantum-dot lasers," *Applied Physics Letters*, vol. 84, no. 10, pp. 1650–1652, 2004. [Online]. Available: <https://doi.org/10.1063/1.1668330>
- [50] C. Ribbat, R. L. Sellin, I. Kaiander, F. Hopfer, N. N. Ledentsov, D. Bimberg, A. R. Kovsh, V. M. Ustinov, A. E. Zhukov, and M. V. Maximov, "Complete suppression of filamentation and superior beam quality in quantum-dot lasers," *Applied Physics Letters*, vol. 82, no. 6, pp. 952–954, 2003. [Online]. Available: <https://doi.org/10.1063/1.1533841>
- [51] T. Akiyama, N. Hatori, Y. Nakata, H. Ebe, and M. Sugawara, "Pattern-effect-free semiconductor optical amplifier achieved using quantum dots," *Electronics Letters*, vol. 38, pp. 1139–1140(1), September 2002.
- [52] E. W. Bogaart, R. Nötzel, Q. Gong, J. E. M. Haverkort, and J. H. Wolter, "Ultrafast carrier capture at room temperature in InAs/InP quantum dots emitting in the 1.55 μm wavelength region," *Applied Physics Letters*, vol. 86, no. 17, p. 173109, 2005. [Online]. Available: <https://doi.org/10.1063/1.1915527>
- [53] P. Borri, W. Langbein, J. Hvam, F. Heinrichsdorff, M.-H. Mao, and D. Bimberg, "Spectral hole-burning and carrier-heating dynamics in InGaAs quantum-dot amplifiers," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 6, no. 3, pp. 544–551, May 2000.
- [54] J. Mørk and T. W. Berg, "The dynamics of semiconductor optical amplifiers: Modeling and applications," *Opt. Photon. News*, vol. 14, no. 7, pp. 42–48, Jul 2003.
- [55] T. Akiyama, M. Sugawara, and Y. Arakawa, "Quantum-dot semiconductor optical amplifiers," *Proceedings of the IEEE*, vol. 95, no. 9, pp. 1757–1766, 2007.
- [56] K. Akahane, T. Umezawa, A. Matsumoto, Y. Yoshida, and N. Yamamoto, "High temperature operation of quantum dot semiconductor optical amplifier for uncooled 80 Gbps data transmission," in *2020 Conference on Lasers and Electro-Optics (CLEO)*, May 2020, pp. 1–2.
- [57] S. Chen, W. Li, J. Wu, Q. Jiang, M. Tang, S. Shutts, S. N. Elliott, A. Sobiesierski, A. J. Seeds, I. Ross, P. M. Smowton, and H. Liu, "Electrically pumped continuous-wave III-V quantum dot lasers on silicon," *Nature Photonics*, vol. 10, no. 5, pp. 307–311, 2016. [Online]. Available: <https://doi.org/10.1038/nphoton.2016.21>
- [58] D. Jung, J. Norman, M. J. Kennedy, C. Shang, B. Shin, Y. Wan, A. C. Gossard, and J. E. Bowers, "High efficiency low threshold current 1.3 μm InAs quantum dot lasers on on-axis (001) GaP/Si," *Applied Physics Letters*, vol. 111, no. 12, p. 122107, 2017. [Online]. Available: <https://doi.org/10.1063/1.4993226>
- [59] C. Shang, Y. Wan, J. C. Norman, N. Collins, I. MacFarlane, M. Dumont, S. Liu, Q. Li, K. M. Lau, A. C. Gossard, and J. E. Bowers, "Low-threshold epitaxially grown 1.3- μm InAs quantum dot lasers on patterned (001) Si," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 25, no. 6, pp. 1–7, Nov 2019.
- [60] C. Shang, E. Hughes, Y. Wan, M. Dumont, R. Kosciwa, J. Selvidge, R. Herrick, A. C. Gossard, K. Mukherjee, and J. E. Bowers, "High-temperature reliable

- quantum-dot lasers on Si with misfit and threading dislocation filters," *Optica*, vol. 8, no. 5, pp. 749–754, 2021.
- [61] A. Y. Liu, S. Srinivasan, J. Norman, A. C. Gossard, and J. E. Bowers, "Quantum dot lasers for silicon photonics [Invited]," *Photon. Res.*, vol. 3, no. 5, pp. B1–B9, 2015. [Online]. Available: <http://www.osapublishing.org/prj/abstract.cfm?URI=prj-3-5-B1>
- [62] Z. Liu, C. Hantschmann, M. Tang, Y. Lu, J. Park, M. Liao, S. Pan, A. Sanchez, R. Beanland, M. Martin, T. Baron, S. Chen, A. Seeds, R. Penty, I. White, and H. Liu, "Origin of defect tolerance in InAs/GaAs quantum dot lasers grown on silicon," *Journal of Lightwave Technology*, vol. 38, no. 2, pp. 240–248, Jan 2020.
- [63] J. Selvidge, J. Norman, M. E. Salmon, E. T. Hughes, J. E. Bowers, R. Herrick, and K. Mukherjee, "Non-radiative recombination at dislocations in InAs quantum dots grown on silicon," *Applied Physics Letters*, vol. 115, no. 13, p. 131102, 2019. [Online]. Available: <https://doi.org/10.1063/1.5113517>
- [64] K. Mukherjee, J. Selvidge, D. Jung, J. Norman, A. A. Taylor, M. Salmon, A. Y. Liu, J. E. Bowers, and R. W. Herrick, "Recombination-enhanced dislocation climb in InAs quantum dot lasers on silicon," *Journal of Applied Physics*, vol. 128, no. 2, p. 025703, 2020. [Online]. Available: <https://doi.org/10.1063/1.5143606>
- [65] J. Selvidge, J. Norman, E. T. Hughes, C. Shang, D. Jung, A. A. Taylor, M. J. Kennedy, R. Herrick, J. E. Bowers, and K. Mukherjee, "Defect filtering for thermal expansion induced dislocations in III–V lasers on silicon," *Applied Physics Letters*, vol. 117, no. 12, p. 122101, 2020. [Online]. Available: <https://doi.org/10.1063/5.0023378>
- [66] A. Liu and J. Bowers, "Photonic integration with epitaxial III–V on silicon," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. PP, pp. 1–1, 07 2018.
- [67] S. W. C. Larry A. Coldren and M. L. Mašanović, *Gain and Current Relations*. John Wiley Sons, Ltd, 2012, ch. Four, pp. 157–246. [Online]. Available: <https://doi.org/10.1002/9781118148167.ch4>
- [68] A. Sobiesierski and P. Smowton, "Quantum-dot lasers: Physics and applications," in *Reference Module in Materials Science and Materials Engineering*. Elsevier, 2016.
- [69] D. Arsenijević and D. Bimberg, "Quantum-dot lasers for 35 Gbit/s pulse-amplitude modulation and 160 Gbit/s differential quadrature phase-shift keying," in *Semiconductor Lasers and Laser Dynamics VII*, K. Panajotov, M. Sciamanna, A. Valle, and R. Michalzik, Eds., vol. 9892, International Society for Optics and Photonics. SPIE, 2016, pp. 123–132. [Online]. Available: <https://doi.org/10.1117/12.2230758>
- [70] K. Takada, Y. Tanaka, T. Matsumoto, M. Ekawa, H. Z. Song, Y. Nakata, M. Yamaguchi, K. Nishi, T. Yamamoto, M. Sugawara, and Y. Arakawa, "Wide-temperature-range 10.3 Gbit/s operations of 1.3 μm high-density quantum-dot DFB lasers," *Electronics Letters*, vol. 47, pp. 206–208(2), February 2011. [Online]. Available: <https://digital-library.theiet.org/content/journals/10.1049/el.2010.3312>
- [71] J. P. Reithmaier, V. Ivanov, V. Sichkovskyi, C. Gilfert, A. Rippien, N. Felgen, F. Schnabel, K. Fuchs, K. Koshuharov, D. Gready, and G. Eisenstein, "InP-based 1.5 μm quantum dot lasers: Static and dynamic properties," in *2013 IEEE Photonics Conference*, Sep. 2013, pp. 1–2.
- [72] Y. Tanaka, M. Ishida, K. Takada, T. Yamamoto, H. Z. Song, Y. Nakata, M. Yamaguchi, K. Nishi, M. Sugawara, and Y. Arakawa, "25 Gbps direct modulation in 1.3- μm InAs/GaAs high-density quantum dot lasers," in *Conference on Lasers and Electro-Optics 2010*. Optical Society of America, 2010, p. CTuZ1. [Online]. Available: <http://www.osapublishing.org/abstract.cfm?URI=CLEO-2010-CTuZ1>
- [73] A. Abdollahinia, S. Banyoudeh, A. Rippien, F. Schnabel, O. Eyal, I. Cestier, I. Kalifa, E. Mentovich, G. Eisenstein, and J. Reithmaier, "Temperature stability of static and dynamic properties of 1.55 μm quantum dot lasers," *Opt. Express*, vol. 26, no. 5, pp. 6056–6066, Mar 2018. [Online]. Available: <http://www.opticsexpress.org/abstract.cfm?URI=oe-26-5-6056>
- [74] M. Ishida, N. Hatori, T. Akiyama, K. Otsubo, Y.

- Nakata, H. Ebe, M. Sugawara, and Y. Arakawa, "Photon lifetime dependence of modulation efficiency and K factor in 1.3 μm self-assembled InAs/GaAs quantum-dot lasers: Impact of capture time and maximum modal gain on modulation bandwidth," *Applied Physics Letters*, vol. 85, no. 18, pp. 4145–4147, 2004. [Online]. Available: <https://doi.org/10.1063/1.1811789>
- [75] P. P. Baveja, M. Li, D. Wang, Y.-Y. Liang, Y. Chen, D. Macintosh-Dorsey, H. Zhang, and J. Zheng, "Impact of laser dynamics on 56 Gbps PAM-4 modulation of 25G class, 1310 nm, directly modulated lasers," in *2017 IEEE Photonics Conference (IPC)*, Oct 2017, pp. 475–476.
- [76] M. Asada, "Intraband relaxation time in quantum-well lasers," *IEEE Journal of Quantum Electronics*, vol. 25, no. 9, pp. 2019–2026, Sep. 1989.
- [77] C. Minkenberg, R. Krishnaswamy, A. Zilkie, and D. Nelson, "Co-packaged data center optics: Opportunities and challenges," *IET Optoelectronics*, vol. 15, no. 2, pp. 77–91, 2021.
- [78] A. Röhm, B. Lingnau, and K. Lüdge, "Understanding ground-state quenching in quantum-dot lasers," *IEEE Journal of Quantum Electronics*, vol. 51, no. 1, pp. 1–11, Jan 2015.
- [79] D. Thomson, A. Zilkie, J. E. Bowers, T. Komljenovic, G. T. Reed, L. Vivien, D. Marris-Morini, E. Cassan, L. Viro, J.-M. Fédéli, J.-M. Hartmann, J. H. Schmid, D.-X. Xu, F. Boeuf, P. O'Brien, G. Z. Mashanovich, and M. Nedeljkovic, "Roadmap on silicon photonics," *Journal of Optics*, vol. 18, no. 7, p. 073003, Jun 2016.
- [80] J. M. Ramírez, H. Elfaiki, T. Verolet, C. Besancon, A. Gallet, D. Neel, K. Hassan, S. Olivier, C. Jany, S. Malhoutre, K. Gradkowski, P. Morrissey, P. O'Brien, C. Caillaud, N. Vaissiere, J. Decobert, S. Lei, R. Enright, A. Shen, and M. Achouche, "III-V-on-silicon integration: From hybrid devices to heterogeneous photonic integrated circuits," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. PP, pp. 1–1, Sep. 2019.
- [81] J. Zhang, B. Haq, J. O'Callaghan, A. Gocalinska, E. Pelucchi, A. J. Trindade, B. Corbett, G. Morthier, and G. Roelkens, "Transfer-printing-based integration of a III-V-on-silicon distributed feedback laser," *Opt. Express*, vol. 26, no. 7, pp. 8821–8830, Apr 2018. [Online]. Available: <http://www.opticsexpress.org/abstract.cfm?URI=oe-26-7-8821>
- [82] A. J. Zilkie, P. Srinivasan, A. Trita, T. Schrans, G. Yu, J. Byrd, D. A. Nelson, K. Muth, D. Leroese, M. Alalusi, K. Masuda, M. Ziebell, H. Abediasl, J. Drake, G. Miller, H. Nykanen, E. Kho, Y. Liu, H. Liang, H. Yang, F. H. Peters, A. S. Nagra, and A. G. Rickman, "Multi-micron silicon photonics platform for highly manufacturable and versatile photonic integrated circuits," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 25, no. 5, pp. 1–13, 2019.
- [83] T. Komljenovic, D. Huang, P. Pintus, M. A. Tran, M. L. Davenport, and J. E. Bowers, "Photonic integrated circuits using heterogeneous integration on silicon," *Proceedings of the IEEE*, vol. 106, no. 12, pp. 2246–2257, Dec 2018.
- [84] R. Jones, P. Doussiere, J. B. Driscoll, W. Lin, H. Yu, Y. Akulova, T. Komljenovic, and J. E. Bowers, "Heterogeneously integrated InP/silicon photonics: Fabricating fully functional transceivers," *IEEE Nanotechnology Magazine*, vol. 13, no. 2, pp. 17–26, April 2019.
- [85] X. Luo, Y. Cheng, J. Song, T. Liow, Q. J. Wang, and M. Yu, "Wafer-scale dies-transfer bonding technology for hybrid III/V-on-silicon photonic integrated circuit application," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 22, no. 6, pp. 443–454, Nov 2016.
- [86] Y. Du, B. Xu, G. Wang, S. Gu, B. Li, Z. Kong, J. Yu, G. Bai, J. Li, W. Wang, and H. H. Radamson, "Growth of high-quality epitaxy of GaAs on Si with engineered Ge buffer using MOCVD," *Journal of Materials Science: Materials in Electronics*, vol. 32, no. 5, pp. 6425–6437, 2021. [Online]. Available: <https://doi.org/10.1007/s10854-021-05360-4>
- [87] C. Shang, J. Norman, A. C. Gossard, and J. E. Bowers, "GaAs epitaxy on (001) Si: Below $1 \times 10^6 \text{ cm}^{-2}$ dislocation density with 2.4 μm buffer thickness," in *2020 Conference on Lasers and Electro-Optics (CLEO)*, May 2020, pp. 1–2.

- [88] Z.-H. Wang, W.-Q. Wei, Q. Feng, J.-J. Guo, M.-C. Guo, T. Wang, and J.-J. Zhang, "InAs/GaAs quantum dot single-section mode-locked lasers monolithically grown on Si (001) substrate with resonant optical feedback," in *Asia Communications and Photonics Conference/International Conference on Information Photonics and Optical Communications 2020 (ACP/IPOC)*. Optical Society of America, 2020, p. T3D.3. [Online]. Available: <http://opg.optica.org/abstract.cfm?URI=ACPC-2020-T3D.3>
- [89] M. Tang, S. Chen, J. Wu, Q. Jiang, V. G. Dorogan, M. Benamara, Y. I. Mazur, G. J. Salamo, A. Seeds, and H. Liu, "1.3- μm InAs/GaAs quantum-dot lasers monolithically grown on Si substrates using InAlAs/GaAs dislocation filter layers," *Opt. Express*, vol. 22, no. 10, pp. 11 528–11 535, May 2014. [Online]. Available: <http://opg.optica.org/oe/abstract.cfm?URI=oe-22-10-11528>
- [90] Y. Xue, W. Luo, S. Zhu, L. Lin, B. Shi, and K. M. Lau, "1.55 μm electrically pumped continuous wave lasing of quantum dash lasers grown on silicon," *Opt. Express*, vol. 28, no. 12, pp. 18 172–18 179, Jun 2020. [Online]. Available: <http://opg.optica.org/oe/abstract.cfm?URI=oe-28-12-18172>
- [91] Q. Li and K. M. Lau, "Epitaxial growth of highly mismatched III-V materials on (001) silicon for electronics and optoelectronics," *Progress in Crystal Growth and Characterization of Materials*, vol. 63, no. 4, pp. 105–120, 2017.
- [92] J. Selvidge, E. T. Hughes, J. C. Norman, C. Shang, M. J. Kennedy, M. Dumont, A. M. Netherton, Z. Zhang, R. W. Herrick, J. E. Bowers, and K. Mukherjee, "Reduced dislocation growth leads to long lifetime InAs quantum dot lasers on silicon at high temperatures," *Applied Physics Letters*, vol. 118, no. 19, p. 192101, 2021. [Online]. Available: <https://doi.org/10.1063/5.0052316>
- [93] M. Sugawara, *Self-Assembled InGaAs/GaAs Quantum Dots*, 1st ed., ser. Semiconductors and semimetals, E. W. Robert Willardson, Ed. Materials Science, Mar. 1999, vol. 60.
- [94] W. W. Chow and F. Jahnke, "On the physics of semiconductor quantum dots for applications in lasers and quantum optics," *Progress in Quantum Electronics*, vol. 37, no. 3, pp. 109–184, 2013.
- [95] M. Gioannini, "Ground-state power quenching in two-state lasing quantum dot lasers," *Journal of Applied Physics*, vol. 111, no. 4, p. 043108, 2012. [Online]. Available: <https://doi.org/10.1063/1.3682574>
- [96] E. E. Vdovin, A. Levin, A. Patanè, L. Eaves, P. C. Main, Y. N. Khanin, Y. V. Dubrovskii, M. Henini, and G. Hill, "Imaging the electron wave function in self-assembled quantum dots," *Science*, vol. 290, no. 5489, pp. 122–124, 2000. [Online]. Available: <https://www.science.org/doi/abs/10.1126/science.290.5489.122>
- [97] T. Nakaoka, J. Tatebayashi, Y. Arakawa, and T. Saito, "Carrier relaxation in closely stacked InAs quantum dots," *Journal of Applied Physics*, vol. 96, no. 1, pp. 150–154, 2004. [Online]. Available: <https://doi.org/10.1063/1.1755857>
- [98] K. Kurata, L. Giorgi, F. Cavaliere, L. O'Faolain, S. A. Schulz, K. Nishiyama, Y. Hagihara, K. Yashiki, T. Muto, S. Kobayashi, M. Kuwata, and R. Pitwon, "Silicon photonic micro-transceivers for beyond 5G environments," *Applied Sciences*, vol. 11, no. 22, 2021. [Online]. Available: <https://www.mdpi.com/2076-3417/11/22/10955>
- [99] B. Shi, B. Song, A. A. Taylor, S. S. Brunelli, and J. Klamkin, "Selective area heteroepitaxy of low dislocation density antiphase boundary free GaAs microridges on flat-bottom (001) Si for integrated silicon photonics," *Applied Physics Letters*, vol. 118, no. 12, p. 122106, 2021. [Online]. Available: <https://doi.org/10.1063/5.0043027>
- [100] GOULDING, M. R., "The selective epitaxial growth of silicon," *J. Phys. IV France*, vol. 02, no. C2, pp. C2-745–C2-778, 1991. [Online]. Available: <https://doi.org/10.1051/jp4:1991290>
- [101] E. Gil-Lafon, J. Napierala, D. Castelluci, A. Pimpinelli, R. Cadoret, and B. Gérard, "Selective growth of GaAs by HVPE: Keys for accurate control of the growth morphologies," *Journal of Crystal Growth*, vol. 222, no. 3, pp. 482–496, 2001.
- [102] D. Guimard, M. Ishida, D. Bordel, L. Li, M. Nishioka, Y. Tanaka, M. Ekawa, H. Sudo, T. Yamamoto,

H. Kondo, M. Sugawara, and Y. Arakawa, "Ground state lasing at 1.30 microm from InAs/GaAs quantum dot lasers grown by metal-organic chemical vapor deposition," *Nanotechnology*, vol. 21, p. 105604, Mar 2010.



Low-Voltage Silicon Photonic Micro-Ring Modulator Design and Stabilization

Dominic J. Goodwill, Patrick Dumais, Jia Jiang, Mohammad M. Rad, Dritan Celso,
Jared C. Mikkelsen, Zehua Li, Eric Bernier *

Abstract

The evolution of high performance ASICs for artificial intelligence, telecommunication, and computer systems is hitting a power consumption wall. At 100 Tb/s throughput, the electrical I/O consumes more power than the processing functions do. Co-packaged optics is smaller and more energy efficient than faceplate-pluggable optics, and is therefore the key to future ASICs. Co-packaged optics requires: 1) high-power shared laser sources or high-density integrated sources; 2) high-density, low-power, and low-voltage DSP chips; 3) specially designed low-voltage CMOS drivers and receivers; 4) high-density modulators; and 5) high-density fiber-to-chip coupling technology. Silicon photonics is the highest density and cheapest wafer for optical integration. In this paper, we pay special attention to micro-ring modulators (MRMs), a key optical modulator technology. Despite extensive research into MRMs, they are yet to be deployed in volume, and foundries have only recently met the tight MRM fabrication requirements. We have recently designed, fabricated, and tested an MRM that features 40 GHz modulation bandwidth, 5 dB modulation loss, and 1.0 V drive for CMOS compatibility. Given that MRMs require precise control of their resonance wavelength, we have also developed heater control technologies and resonance training algorithms to ensure error free transmission.

Keywords

silicon photonics, optical interconnect, micro-ring modulator

* Corresponding author.

1 Introduction

Datacenters, compute clusters, and wireless network cores employ high-performance CMOS ASICs interconnected together to process or route information. System capacity requirements continue to increase, and cooling capabilities have already become a limiting factor [1]. As such, it is vitally important to decrease the energy needed to connect and process each bit. In recent ASIC generations, two-thirds of the electrical power is used for core processing functions, while the remaining third is used for off-chip interfaces [2]. However, the future trend is not favorable. Although Moore's Law continues to drive down the energy per bit of ASIC processing cores, the energy efficiency of electrical interfaces is not keeping pace. Without a change in the physical architecture, the interfaces will soon consume the vast majority of the available power. Clearly, this is not a sustainable trend, as capacity growth would stall.

We are motivated to reverse this trend. In this paper, we present our high-density silicon photonics optical interconnect technology, which will allow ASIC interfaces to achieve higher energy efficiency.

1.1 Optical Interconnect Application Space

The application space for optical interconnects continues to grow. Currently, the IEEE 802.3 Ethernet Working Group is developing standards for a single-channel 200 Gb/s interconnect capability with 800 Gb/s per interface [3]. These objectives include a wide variety of optical interconnects, spanning from meters to tens of kilometers. The objectives also include electrical cables, but with a maximum reach of only 1 m due to signal degradation at high signaling rates. This 1 m limit is 50% lower than the previous generation and insufficient even to connect a full equipment rack. Consequently, as the data rate increases, we predict that optical interconnects will continue to penetrate the cabling market to ever-shorter distances.

1.2 Current Architecture: Faceplate Pluggable

Today's systems consist of ASICs assembled onto low-cost PCBs, which are installed into systems with an edge of the PCB at a mechanical faceplate. Optical transceiver modules



Figure 1 Huawei CloudEngine with 400GE QSFP-DD faceplate pluggable optical interconnects [4]

are plugged through the faceplate, into electro-mechanical cages on the PCB, to provide optical I/O from the PCB. An example is shown in Figure 1. Each pluggable form factor typically has multiple interchangeable standard variants.

The faceplate pluggable architecture, which has been ubiquitous for 20 years, has two key operational benefits.

1. **Non-homogeneity:** Operators can mix and match different interconnect modules depending on the distance between individual systems and their choice of suppliers.
2. **Field replaceability:** Operators can unplug and replace an interconnect module if it fails, with minimal skill and no tools. For clarity, in the traditional architecture, replacing a failed interconnect module does not require replacing an entire card.

1.3 Excessive SerDes Power

Despite its benefits, the faceplate pluggable architecture comes with a major drawback: it wastes electrical power. Each data path involves three concatenated interconnects:

- ASIC #1 to faceplate #1 through tens of centimeters of circuit board. (Electrical interconnect #1)
- Faceplate #1 to faceplate #2 for system-to-system cabling. (Optical interconnect)
- Faceplate #2 to ASIC #2. (Electrical interconnect #2)

At present-generation signal rates, the circuit boards create electrical attenuation reaching tens of dB. This degrades the signal so much that strong electrical recovery and error correction SerDes are needed at both ends of the link between the ASIC and faceplate pluggable module [5]. Such a SerDes is known as very short reach (VSR).

For a 50 Tb/s switch using 56 Gbaud SerDes, the SerDes on the ASIC uses ~40% of the ASIC power budget [2]. Until now, the operational benefits of the faceplate pluggable architecture have been more important than this high power consumption. In the next few years, as the external connectivity of an ASIC moves beyond 100 Tb/s, the SerDes will consume a growing proportion of the power budget and will therefore become a formidable problem.

As an example, a VSR electrical SerDes at 100 Gb/s uses around 3 pJ to transmit and receive each bit of data. For an ASIC that transmits 100 Tb/s of data, 3 pJ/bit is equivalent to 300 W. In the largest ASICs today, conventional forced-air cooling can handle up to 400 W [1]. This appears to be a hard limit, as the capacity of forced-air cooling has increased by only 30% over the last 10 years. Consequently, for a 100 Tb/s ASIC, 75% of the conventional cooling budget would be assigned to its interfaces, leaving only 25% for the ASIC core. This is not an attractive solution.

In the short term, other more exotic cooling technologies can be used to mitigate this problem [1]. However, there will come a point in time when the power consumption of SerDes needs to decrease [2].

1.4 Reducing the SerDes Power with Co-packaging

Silicon photonics (SiPh) allows for the insertion of optical interconnects very close to the ASIC, eliminating the need for signals to traverse tens of centimeters of lossy circuit board. This approach not only reduces signal degradation significantly, but also requires fewer pJ/bit for SerDes. Our industry is standardizing [5, 6] several packaging arrangements for socketed or soldered near-package optics (NPO) and co-packaged optics (CPO), as well as several lower-power SerDes such as the digital-to-digital extra short reach (XSR) family [6]. An illustration of one use-case is shown in Figure 2. The optical engine may be a chiplet that embeds all the optical I/O functions into a small die. The industry is well aware that these new architectures affect reliability and non-homogeneity.

The energy-per-bit involved in SerDes may be reduced even further with a massively parallel, low baud-rate scheme. Such a scheme could adopt a 10 to 25 Gbaud signaling rate, the NRZ coding scheme, and lower-gain forward error correction (FEC), or no FEC in shorter-distance applications, while employing significantly more lanes [7, 8]. Using such

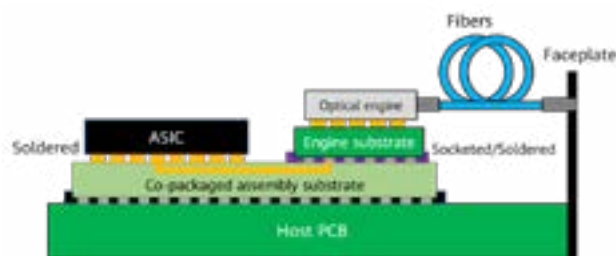


Figure 2 Co-packaged optics: example use-case

a simple SerDes and no FEC minimizes latency. Low latency is particularly advantageous for compute applications. The optical engines may also be 3D assembled directly with the ASIC to reduce not only the electrical parasitics, but also the SerDes performance requirements.

We refer to these techniques collectively as "Optical I/O", because they replace the high-performance "Electrical I/O" of today's ASICs.

1.5 Silicon Photonics Optical I/O

We define SiPh as silicon-on-insulator photonics. It contains active electro-optics (modulators) using a silicon optical waveguide layer, with on-chip photodetectors, optical routing, and fiber couplers. At some foundries, transistors, lasers, semiconductor optical amplifiers, and electro-absorption modulators can also be made on the same chip. SiPh dies are manufactured on existing CMOS production equipment and tested at wafer scale.

Using SiPh, we can address the operational concerns brought by shifting away from the faceplate pluggable architecture.

- **Homogeneity:** When manufactured at high volume, SiPh is inexpensive, and manufacturing costs are only weakly related to circuit complexity. As such, a one-size-fits-most design can be cost-effective, addressing the homogeneity problem.
- **Reliability:** The reliability of SiPh dies is on par with the ASIC, achieving CMOS-like reliability. In an optical device, lasers are the least reliable element. To address this, SiPh enables complex optical circuits to be realized at minimal additional manufacturing cost, allowing redundant lasers to be placed on the SiPh chip, which can then detect failures and promptly switch to the backup laser by means of a (reliable) on-chip 2 x 1 photonic switch element [9]. Further, recent reports show that lasers on SiPh may be much more reliable than standalone lasers, as several root causes of laser

failure are avoided [10]. Alternatively, external laser sources can be used in faceplate pluggable modules [1], in which case the SiPh chip provides sophisticated connectivity and monitoring of the external lasers.

Both these techniques significantly improve the reliability of an optical I/O. Reference [1] sets a target of 10 FIT for a 3.2 Tb/s optical transceiver engine (intended for use as a group of 16 engines with a 50 Tb/s ASIC). However, it allows a more relaxed 50 FIT for an optional external laser source because such a laser is field-replaceable.

Our SiPh solution is compatible with both light source techniques, which are both likely to exist in the market. While the on-chip laser has lower in-system optical loss, the external laser is housed in a lower temperature part of the system, facilitating higher laser efficiency.

1.6 Low-Voltage Modulators Enable Optical I/O

Even when the SerDes is simplified, the optical link still requires its own sophisticated transmit and receive electronics to operate at 100 Gb/s/lane and higher. In addition to implementing SerDes for the electrical interface side, these electronics also provide optical transmitter and signal recovery functions such as equalizers and FEC. These functions comprise complex mixed-signal and digital circuits. The industry trend is to implement such functions entirely within the Tx/Rx functionality using small node CMOS. This approach minimizes the power consumption, particularly in the DSP elements of the transceiver.

However, a limitation of small node CMOS is that the maximum output voltage swing is only around 1 V. As such, designs for low voltage swing are most appropriate for SiPh modulators, as we do not wish to add discrete Tx driver electrical amplifier die which are challenging to fit within a small co-packaged footprint. We describe our low-voltage SiPh modulators later in this paper.

1.7 MRMs Enable Smaller Optical Engines to Minimize SerDes Requirements

To minimize the electrical loss and consequently minimize SerDes power consumption, the optical engine needs to be located as close as possible to the ASIC. It also requires

at least 32 lanes, and possibly more, to achieve sufficient throughput (~Terabit per cm of package perimeter beachfront) at a practical lane rate (up to 200 Gb/s). Because space is tight, each optical lane needs to occupy a very small footprint [11].

For the optical receivers, space is not a problem, as optical receivers are already ultra-compact. All high-speed photodetectors, including SiPh photodetectors, are a fraction of the size of a solder microbump (which is $\leq 60 \mu\text{m}$ radius). As such, the optical part of the Rx is very small. Using on-chip optical routing, we place each photodetector directly underneath its corresponding electrical Rx circuit to achieve the best signal integrity. Rx density is then limited by the CMOS footprint.

For the optical Tx, space is governed by the size of the optical modulator. A typical modulator in commercial use today is the Mach-Zehnder modulator (MZM), which is relatively large. Existing SiPh MZMs are typically millimeters long and have a pitch of at least $250 \mu\text{m}$. Thin-film lithium niobate (TFLN) MZMs are currently in development. Reference [12] reports a TFLN MZM that is 5 mm long by $380 \mu\text{m}$ wide using a 3.6 V drive swing. Reference [13] reports a TFLN MZM that is 7 mm long by several hundred micrometers wide using a 1.15 V drive swing. A total of 32 MZM lanes would occupy a width of about 14 mm, including packaging. While this would fit within a first-generation OIF CPO optical engine [5], shrinking it further for 100 Tb/s ASIC CPO would be challenging. Figure 3 shows a possible floorplan for CPO modules in a 100 Tb/s ASIC solution.

We have therefore developed optical MRMs for the optical Tx. An MRM has an active area radius of $10 \mu\text{m}$ or less, occupying less space than a solder microbump [7].

When deployed in an optical bus configuration, each MRM can address a different wavelength, leaving the other

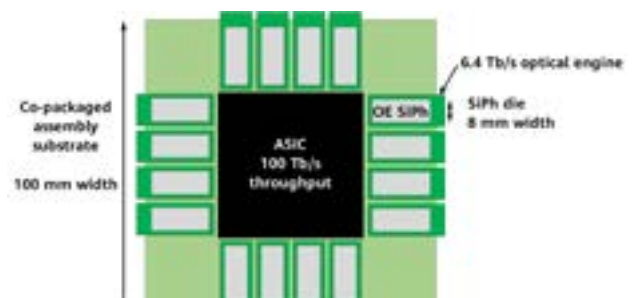


Figure 3 Floorplan of 100 Tb/s ASIC co-packaging with 16 optical engines, assuming 200 Gb/s/lane. Electrical trace length must be minimized, resulting in a very narrow CPO package width.

wavelengths on the bus unaffected. It is therefore possible to easily implement a WDM transceiver in the SiPh chip.

As optical I/O capacity requirements increase, the footprint per lane must shrink. At the same time, the number of wavelengths must increase and the energy per bit must decrease. Looking ahead to the future, the benefits of low-voltage SiPh MRMs will become more and more compelling.

2 Modulator Design and Simulations

2.1 Overview

The design of an MRM, shown conceptually in Figure 4, is a complex multi-physics problem that must be modeled using optical, electrical RF, thermal, and TCAD simulations. The performance optimization to be carried out involves multiple variables and objectives, often requiring several design iterations to converge to the optimal MRM design in the overall system context. Once parameters such as the drive voltage swing and the target baud rate are determined, the phase shifter cross-section, dopant level, and coupler value can be determined in order to optimize the optical modulation amplitude (OMA) or an equivalent metric of modulation effectiveness. During performance evaluation, consideration should be given to the parasitic resistances and capacitances of metal traces and other physical elements.

2.2 Phase Shifter

In our MRM designs, the phase shifters driving the modulation comprise reverse-biased p-n junctions. In this operating mode, also known as carrier-depletion mode, a change in electrical bias changes the width of the depletion region, which in turn changes the complex effective index of

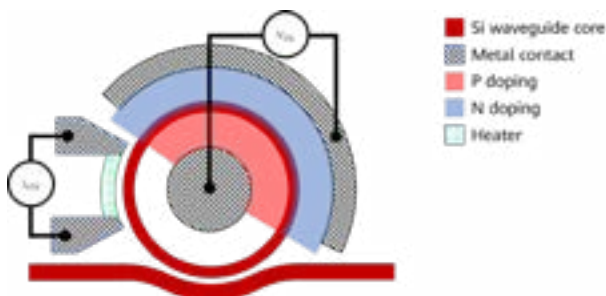


Figure 4 MRM conceptual layout

refraction of the MRM optical waveguide mode. The result is an optical phase shift that slightly changes the resonance and causes a change in the ring transmission. Figure 5a illustrates the transmission curve characteristic to the MRM and the change in that curve resulting from a change in voltage on the p-n junction. Depletion-type modulators can reach speeds in excess of 90 Gbaud [14, 15], making them ideal candidates for 100G/200G per lane transceiver applications.

Lateral p-n (LPN) designs, in which the left and right sides of the waveguide core are oppositely doped, result in a depletion region that extends vertically through the waveguide. While this design simplifies the doping window placement and the junction fabrication, it tends to be a sub-optimal solution due to poor overlap of the depletion region with the optical mode. For this reason, tailoring the doping recipe and window placement to form specialty junctions featuring much higher overlap with the optical mode has received a great deal of attention [16]. Because the required implant doping is highly inhomogeneous in the vertical direction, we refer to these designs broadly as vertical p-n (VPN) designs, and characterize them as U-shaped,

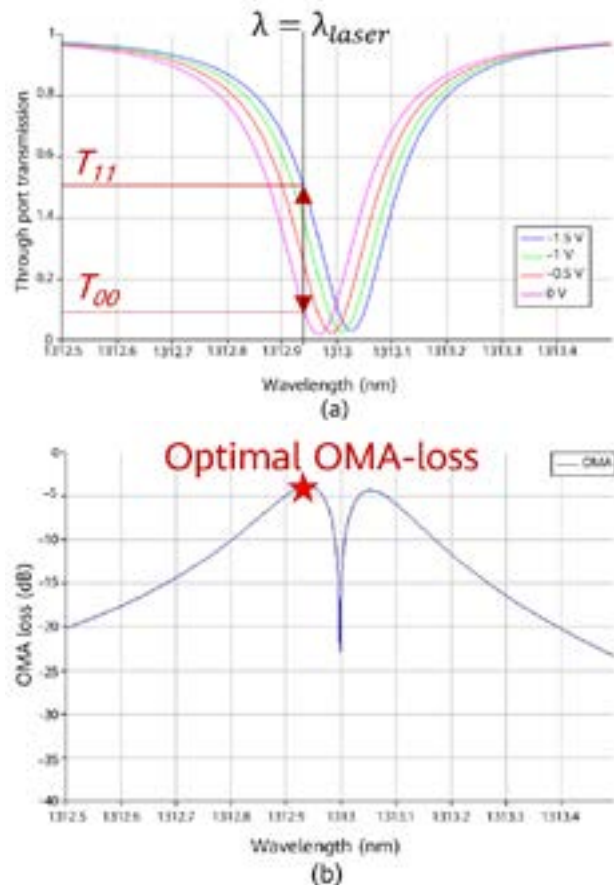


Figure 5 (a) Normalized MRM spectra as a function of applied voltage, showing a shift of the resonance as the reverse bias is increased; (b) OMA-loss as a function of laser detuning, assuming that the voltage across the junction is driven from -1.5 V to 0 V during modulation

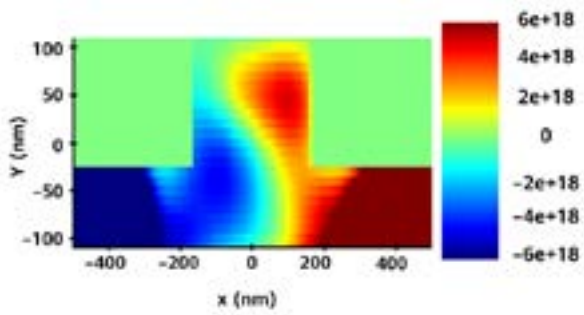


Figure 6 Simulated net doping concentration per cm³ for O-band VPN-doped MRM. The junction has a curved shape.

S-shaped, or L-shaped junctions depending on the shape of the depletion region [17].

For these preferentially shaped junctions, the depletion region generally extends horizontally across the waveguide core. As the waveguide core is typically wider than it is high, assuming a 220 nm thick Si platform, a larger overlap between the depletion region and the waveguide mode occurs than for LPN junctions. VPN junctions therefore achieve a higher modulation efficiency (characterized as the index change per applied voltage). Furthermore, because their junction area is larger, VPN junctions have higher capacitance per unit length than LPN junctions do. This higher capacitance reduces the RC frequency limit of VPN junctions to some extent.

By tailoring the implant energies and doses appropriately, a "top-heavy" and "bottom-heavy" implantation for each dopant species respectively can be combined to form a VPN junction where they overlap. This approach is used in our MRM design. The simulated doping concentration in the ring waveguide cross-section for an O-band VPN MRM is shown in Figure 6.

2.3 Achieving High-Speed Modulation

The modulation bandwidth of the MRM is determined by the photon lifetime in the ring and the RC electrical bandwidth. The Q-factor is often considered a proxy for the optically-limited modulation speed of the ring because it is related to the photon cavity lifetime τ by:

$$\tau = \frac{Q\lambda_{res}}{2\pi c} \tag{1}$$

Here, τ represents the characteristic 1/e rise (or fall) time of the intra-cavity optical intensity of the micro-ring,

for a CW laser input at the cavity resonant wavelength. The associated 3 dB roll-off frequency of the modulated through-port signal is then:

$$f_{3dB,resonant\ optical} = \frac{1}{2\pi\tau} \tag{2}$$

However, it is important to note that Equation (1) applies only when the laser wavelength is exactly aligned to the micro-ring resonance. This is a lower bound to the maximum frequency. In practical operation, the laser wavelength is somewhat detuned from resonance, and speeds beyond that of Equation (1) are achieved. MRMs generally target a Q-factor in the 3000–7000 range, because these values provide a reasonable trade-off between speed, OMA, and drive voltage swing. For a device optimized for a low-voltage CMOS driver, the Q-factor should be ~6700.

Figure 7 shows the electrical model for the p-n junction. The circuit parameters of the MRM waveguide phase shifter are determined using small-signal analysis, while the junction series resistance and capacitance are calculated from real and imaginary parts of the resulting complex impedance. Figure 8 shows the calculated small-signal resistance R and capacitance C . The resistance is relatively constant, whereas the capacitance decreases in a power-law fashion with increasing reverse bias.

In addition to the MRM phase shifter itself, the parasitic resistances, capacitances, and inductances of the surrounding pads, contacts, and wire traces must be considered. Figure 9a shows the parasitic circuit model used in our MRM designs. We used a combination of geometrical calculations and foundry-provided reference values to determine worst-case estimates of circuit values. The effect of the parasitic circuit on the electrical transfer function (i.e., the junction voltage normalized to the driver source) is shown in Figure 9b. This effect can reduce the roll-off frequency by several GHz. As such, we conclude that the small capacitance increase due to the optimized junction is not a major limiting factor in the device speed.

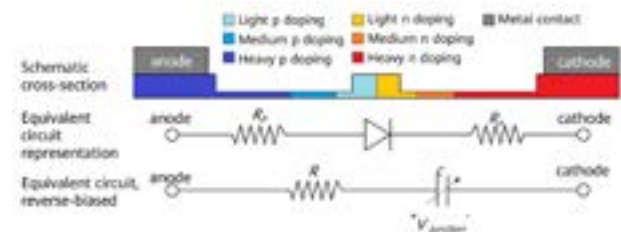


Figure 7 Electrical network representation of the MRM waveguide electro-optic phase shifter

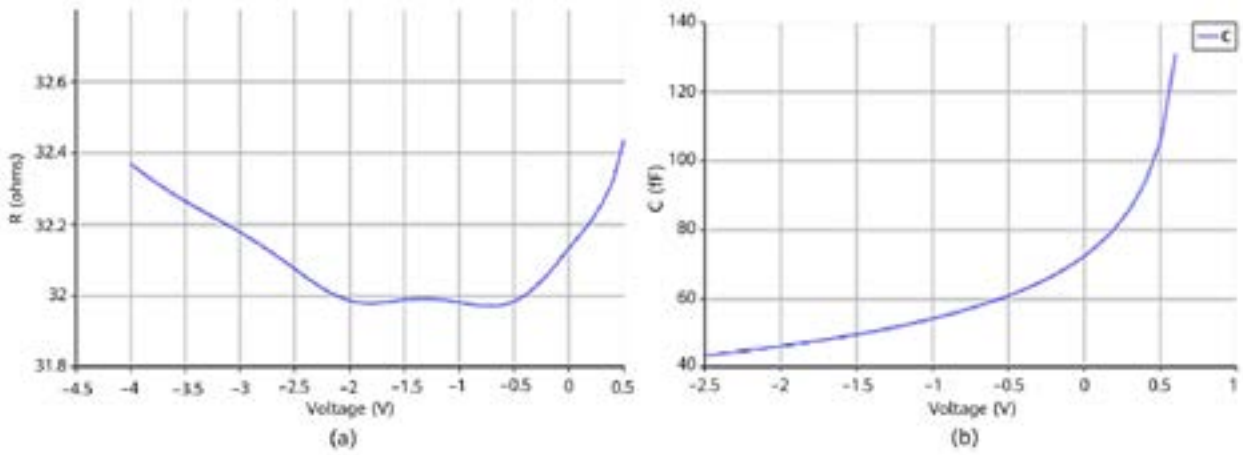


Figure 8 (a) Junction series resistance and (b) junction capacitance of the MRM waveguide phase shifter, as a function of voltage bias. Results were calculated using a small signal at a frequency of 1 GHz.

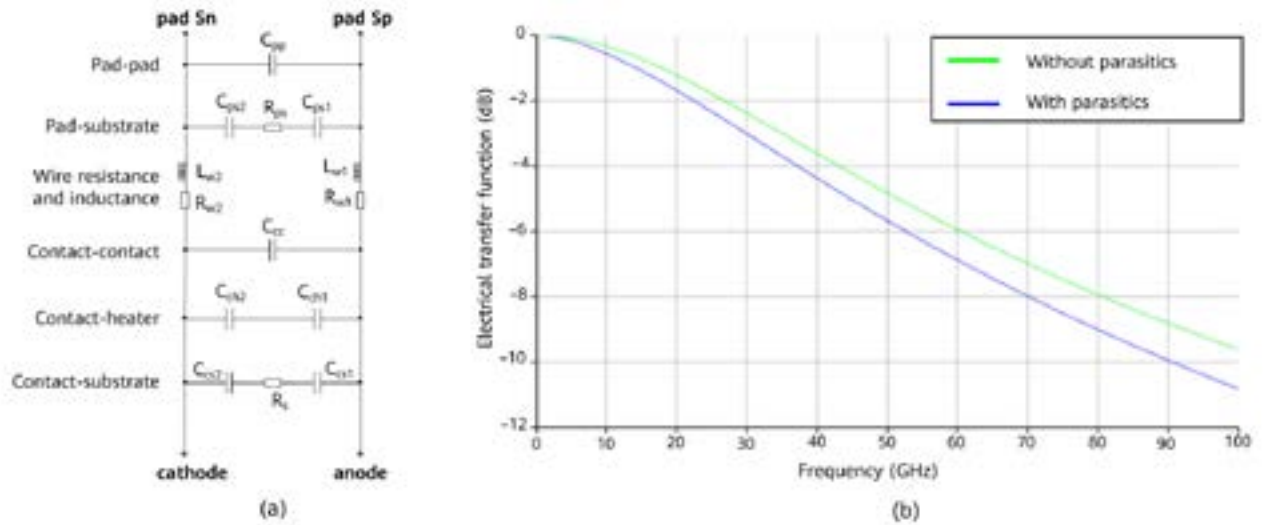


Figure 9 (a) Electrical parasitic circuit schematic for MRM; (b) Effect of worst-case electrical parasitics on the electrical transfer function of the MRM, given a driver source impedance of 30 Ω

2.4 High-Speed Response

In the absence of nonlinear optical effects, a small-signal sinusoidal modulation applied to the electrical contacts of the MRM electro-optic phase shifter produces a corresponding small-signal sinusoidal modulation of the same frequency in the optical amplitude at the output port. The relative amplitude and phase between the output optical amplitude and the input voltage phasor is referred to as the electro-optic S_{21} (EOS_{21}) of the MRM. The EOS_{21} response depends on both the electrical set-point (phase shifter bias) and the optical set-point of the MRM (laser wavelength relative to MRM instantaneous optical resonance wavelength). When the laser wavelength is

aligned with the MRM resonance, the total MRM bandwidth is the combination of the electrical and optical bandwidths:

$$f_{3dB,resonant} = ((2\pi\tau)^2 + (2\pi RC)^2)^{-1/2} \quad (3)$$

For intra-cavity index-based MRM modulation, the laser wavelength cannot be equal to the resonance. This is because the transmission slope at that point is 0, which would result in a severely distorted output signal. On the short-wavelength side of the resonance, the laser wavelength is typically detuned to the wavelength point that provides the maximum OMA value. Such detuning results in a roll-off frequency beyond that predicted by Equation (3). As detuning increases, the EOS_{21} exhibits peaking behavior, similar to that of an inductive peaking amplifier [18].

Figure 10a and Figure 10b show the magnitude and phase response of the EOS_{21} for the designed device for a drive voltage that swings from -1.5 V to 0 V. The driver source impedance is set at 30Ω , approximately matching the simulated junction series resistance. Due to the changing laser detuning, a significant dependence of EOS_{21} peaking behavior and roll-off frequency on voltage is observed. Figure 10c shows the 3 dB roll-off frequency at mid-swing (0.75 V bias) versus OMA-loss, as the laser detuning is swept. OMA-loss is defined as the magnitude of the optical output power swing divided by the optical input power, and is expressed in optical dB. Modulators have a general design goal of maximizing OMA or, equivalently, minimizing OMA-loss.

For the designed device, the mid-swing roll-off frequency is about 37 GHz at the point of maximum OMA-loss. For

detuning values approaching 0 , the roll-off frequency approaches that predicted by Equation (3). By increasing the detuning, the roll-off frequency can be traded off against OMA-loss, as shown in Figure 10c.

2.5 Foundry Requirements

A well-designed modulator must meet the target OMA loss over the expected fabrication variations [19]. In our modulator design, the most demanding fabrication tolerance is for the silicon etch depth. This is because the etch depth affects the power coupling coefficient of the ring-bus optical coupler. Variation of the ring-bus coupling affects OMA-loss and (by changing the photon lifetime) modulation bandwidth. Our modulator design calls for a nominal coupling value of 10% optical power.

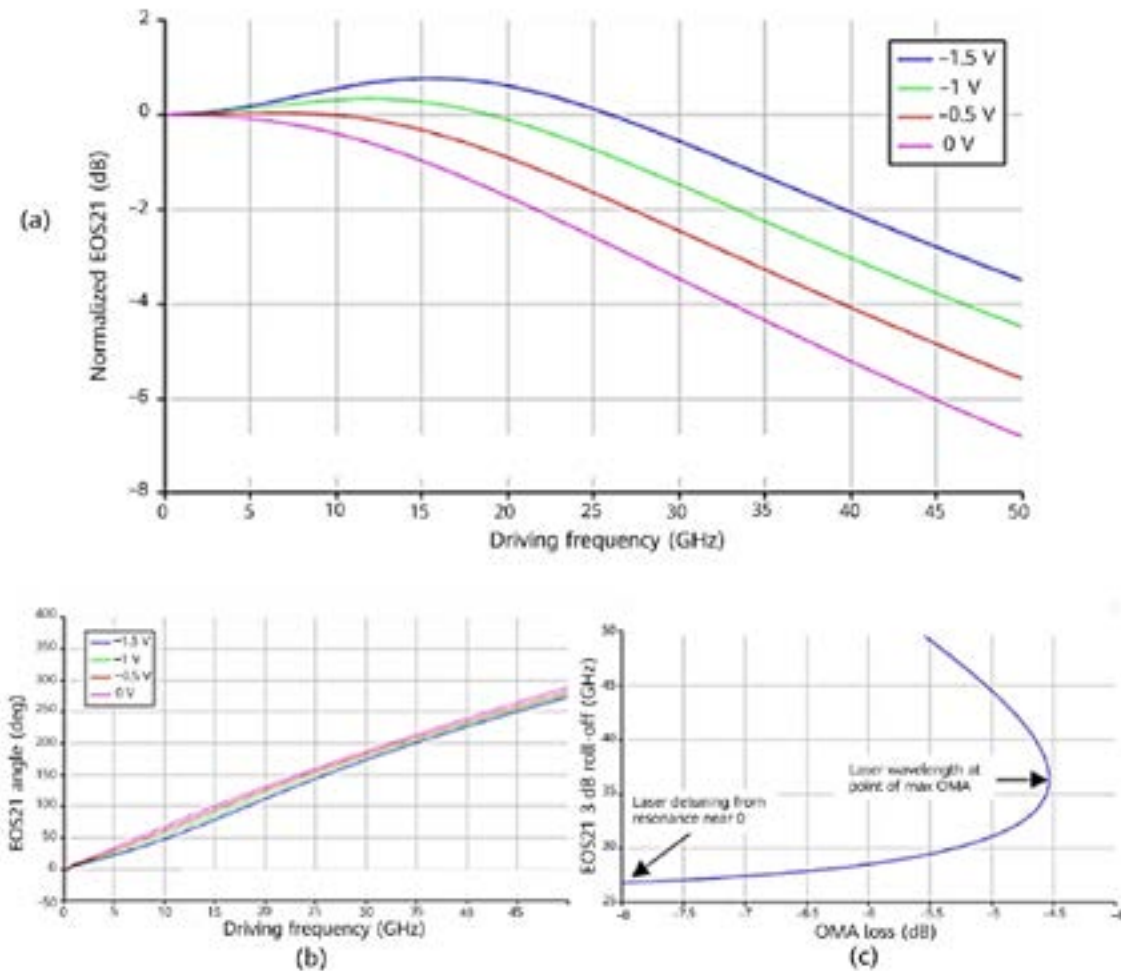


Figure 10 (a) Normalized EOS_{21} magnitude and (b) EOS_{21} angle of MRM as a function of bias voltage. The laser wavelength is set differently at each point on the graph in order to maximize the OMA for a -1.5 V to 0 V swing at that operating point. (c) EOS_{21} 3 dB roll-off frequency at mid-swing, plotted against OMA-loss as the detuning is changed.

We implement the ring-bus coupler as a curved ring coupler using silicon waveguides with partially-etched shoulders. A curved coupler, shown in Figure 11a, allows a longer coupler region and a higher tolerance to process parameter variation. The waveguide in the active part of the ring also has partially etched silicon shoulders that are doped to provide electrical contacts. Therefore, using partially etched waveguide shoulders in the coupler permits a compact layout, as we do not need a waveguide-type transition region between the coupler and the active part of the ring. This is in contrast to designs that use couplers with fully-etched shoulders.

The waveguide width variation, silicon thickness variation, and etch depth variation are the three fabrication process parameters that affect the coupler value. Figure 11b shows the spread of coupler values when varying these parameters over the surface of an ellipsoid representing a combined 3σ variation of the parameters. The 3σ variations for the individual parameters were ± 8 nm, ± 5 nm, and ± 14 nm for the width, thickness, and etch depth respectively. Figure 11b

shows that the power coupling coefficient is almost linear with the remaining slab thickness after etch, demonstrating that this is the most important parameter affecting the coupler value.

We estimate that an acceptable range of variation is $\pm 2\%$ [19], which meets the target bandwidth while preserving the OMA loss within 1 dB of nominal. This in turn dictates that the etch depth tolerance should be restricted to ± 9 nm, keeping the other tolerances as-is. We believe this level of uniformity can be achieved only in commercial foundries that use 300 mm wafers.

In Table 1, we list the foundry process requirements for MRM, such as the coupler requirements, and present typical capabilities for 300 mm foundries, 200 mm foundries, and typical e-beam test-device fabrication. A foundry capable of producing 300 mm wafers meets etch depth uniformity requirements. The 45–55 nm nodes have an adequate dopant overlay tolerance, allowing the junction shape to be adequately controlled and the dopant levels to be precisely placed. This ensures that the MRM bandwidth is not RC-limited.

Table 1 Foundry process requirements for MRM

Process item	Requirement	45–55 nm Node 300 mm	130 nm Node 200 mm	E-Beam 200 mm
Etch variation	± 9 nm	± 7 nm	± 10 nm	No info
Overlay tolerance	± 15 nm	± 10 nm to ± 20 nm	$< \pm 50$ nm	No. Contact doping ± 1 μ m
UBM	Must have	Yes	?	Not typical
TSV	Should have	?	?	Not typical
Edge coupler	Should be on foundry roadmap	Yes	Depends on foundry	Not typical
Number of metal layers	At least 2 metal layers	Yes	Depends on foundry	Only 1 typical
Heater metal composition	High temperature tolerant (250 °C) TiN, W, or other	?	Depends on foundry	Not typical
Doping profiles	One of: Custom doping allowed Standard doping allows vertical PN junctions	?	Yes	Yes
Passive waveguide loss	< 2 dB/cm loss for rib waveguide	Yes	Yes	Not guaranteed
Feature size variation	± 15 nm	$< \pm 10$ nm	$< \pm 10$ nm	?
Ge PD	Must have	Yes	Yes	Not typical

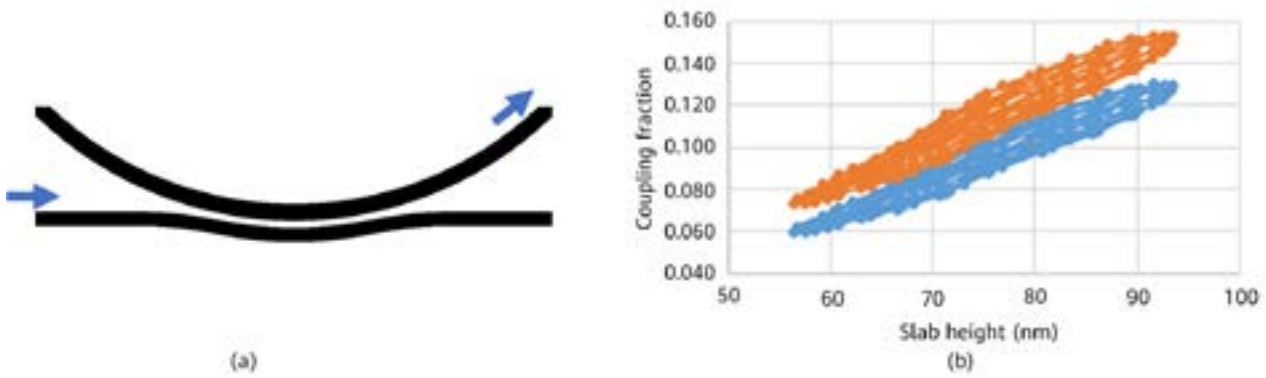


Figure 11 (a) Layout of curved ring-bus coupler; (b) Wavelength-dependent variation of the coupling value over the 3σ ellipsoid of three parameters (see text). Blue indicates the shortest wavelength, and orange indicates the longest wavelength.

3 Experimental Results of MRM Testing

3.1 MRM Test Chip Fabrication

We fabricated MRMs designed for O-band operation and MRMs designed for C-band operation. The MRM radius in all cases was $10.0\ \mu\text{m}$. A photograph of typical fabricated devices is shown in Figure 12.

The layer stack included $2\ \mu\text{m}$ buried oxide, $220\ \text{nm}$ crystalline silicon core, and $1\ \mu\text{m}$ silica top cladding layer. Ring and bus waveguides were defined using a $130\ \text{nm}$ etch, leaving a $90\ \text{nm}$ silicon slab throughout the devices. Routing waveguides connected Si grating couplers to the modulators, and were fully-etched silicon features.

The modulation region of the MRM comprised a vertically-doped p-n junction and p⁺⁺/n⁺⁺ contacts. N-doped heaters were used to tune the MRM resonance, and were electrically

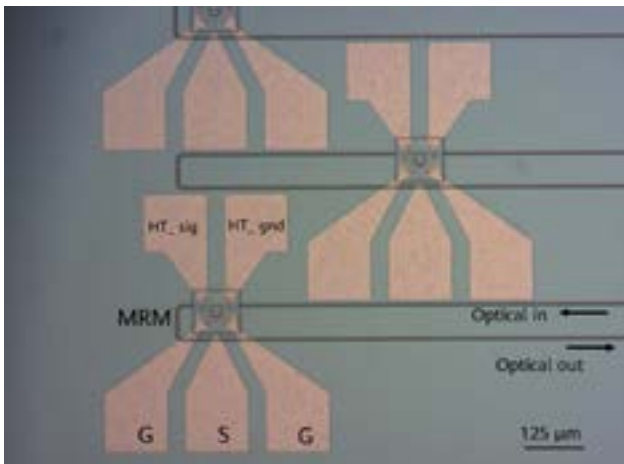


Figure 12 Photo of an MRM test die

isolated from the p-n junction. All electrical traces and contact pads comprised a single layer of aluminum.

3.2 Test Configuration

The MRM test configuration is shown in Figure 13. CW light from a tunable laser was injected through a surface grating coupler array using an angled 10-fiber array unit (FAU) with a fiber pitch of $127\ \mu\text{m}$. RF signals and DC bias were injected to the MRM using a $67\ \text{GHz}$ GSG RF probe and DC probes. Output light was fiber-coupled from an output grating coupler to a high-speed photoreceiver.

The electro-optic performance of the MRM was measured using a programmable vector network analyzer (PNA) (Keysight N5225A) with $50\ \text{GHz}$ bandwidth. The DC bias voltages were generated by voltage reference equipment and applied through an external bias-T.

The response of the probes, RF cables, Bias-T, connectors, and photoreceiver were de-embedded by manufacturer-recommended calibration techniques, for accurate EOS_{21} measurements.

3.3 Optical Passive Testing

Figure 14 shows a typical measured transmission spectrum of our C-band MRM. The average free spectral range (FSR) is $6.6\ \text{nm}$ and $9.7\ \text{nm}$ at $1310\ \text{nm}$ and $1550\ \text{nm}$ respectively, for all the measured devices. This is within $\pm 1\%$ of the design value. The quality factors (Q) were calculated to be ~ 3000 , which is slightly lower than expected (~ 3800 for this chip design), likely due to higher than expected propagation loss within the MRM. The extinction ratio at resonance was

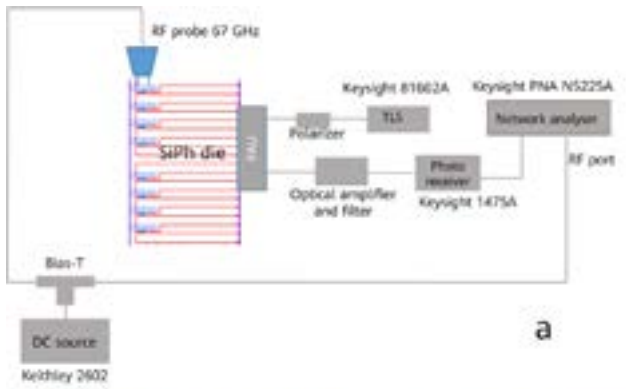


Figure 13 Testing setup. (a) Equipment diagram; (b) RF probe, fiber ribbon, and chip on the testing station; (c) Close-up view of the RF probe, fiber ribbon, and MRM die

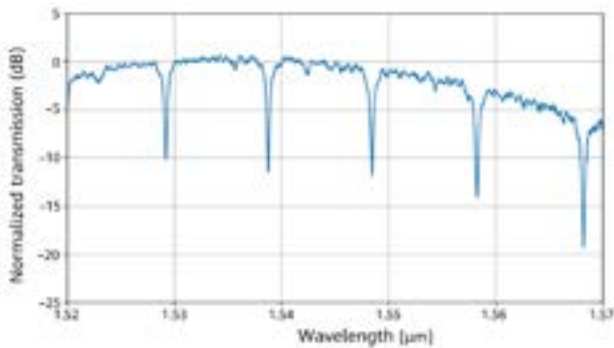


Figure 14 Measured transmission spectrum of C-band MRM. Resonances may be seen as notches spaced by 9.7 nm FSR. Modulation tests were performed on the short-wavelength side of the 1538 nm resonance.

8 to 15 dB. The ring-bus coupling ratio extracted from the spectrum of Figure 14 is within a 7% to 15% interval, which is consistent with the nominally-designed coupling ratio of 10%. The spectrum shows ripples, which we attribute to surface grating coupler reflections, and the resonances are asymmetric, which we attribute to optical self-heating within the MRM.

3.4 DC Electro-optic Test

The junction resistance was determined from the DC I-V curve as $0.74 \Omega \cdot \text{cm}$ to $1.06 \Omega \cdot \text{cm}$, which is consistent with our calculation using the expected dopant levels. This resistance will be lowered in a production foundry process, by adding intermediate-doped regions between the junction and the contacts. The EOS_{21} bandwidth will then improve further.

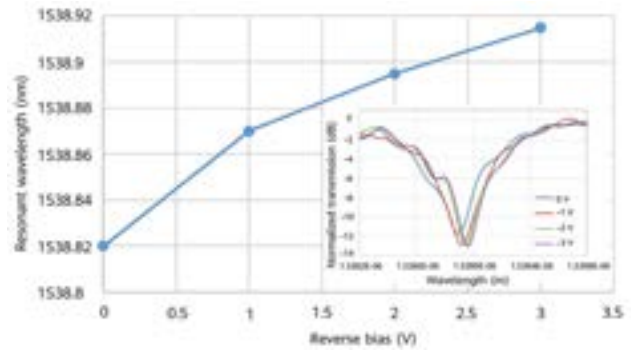


Figure 15 MRM modulation efficiency, DC. Inset: Spectra of an MRM resonance versus applied DC voltage.

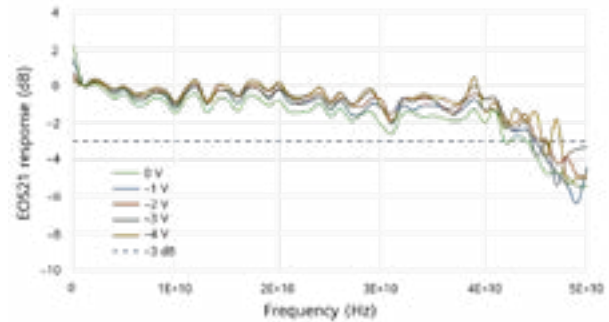


Figure 16 Measured small-signal EOS_{21} of MRM at bias voltages from 0 V to -4 V. Dashed line indicates 3 dB roll-off.

A variable DC voltage was applied and the change in optical spectrum was measured. The typical resonance peak shift induced by an applied DC voltage is shown in Figure 15. The ring resonance shifts by 65 pm/V averaged over a voltage range of 0 V to -1 V regime. The ring resonance shifts by 43 pm/V averaged over a voltage range of 0 V to -3 V.

The average $V\pi L$ measured over multiple MRM devices on the same chip was $0.69 \text{ V} \cdot \text{cm}$ (measured range was 0.5 to $0.93 \text{ V} \cdot \text{cm}$). The measured $V\pi L$ is consistent with a well-defined VPN junction having p-n doping concentrations $\sim 6 \times 10^{18} / \text{cm}^3$, which is consistent with the design.

3.5 High-Speed Electro-optic Test

The small-signal RF electro-optic response (EOS_{21}) was measured at DC bias of 0 V to -4 V, and is shown in Figure 16. The MRM bias point was 3 dB from the resonant wavelength. As shown in Figure 15, the 3 dB bias point corresponds to a wavelength detuning of 0.43 nm (blue-shifted) away from the bottom of the resonance dip. The EOS_{21} 3 dB optical bandwidth for the C-band MRMs was measured to be greater than 40 GHz for all bias voltage levels, with a maximum of 45 GHz at -3 V bias voltage.

40 GHz is sufficient for at least 64 Gbaud modulation without excessive electrical equalization requirements.

4 MRM Control

4.1 Resonance Sensitivity for MRM

Silicon modulators are predominantly of two types: MZM and MRM.

The MZM is a balanced device (two identical arms) with nominally zero-order wavelength resonance. Simply put, the optical path difference between the two arms of the device is close to zero. The Q-factor of an MZM is 1, because the MZM is not optically resonant. Given that an MZM is essentially a white-light device, after initial correction of the optical phase, it is relatively immune to environmental temperature changes, drift over life, drifts of laser wavelength, and incident RF and optical power.

However, to shrink the modulator into a very small area while retaining the same level of performance, the modulator must be rolled up into a resonant device, to form an MRM. MRMs are single-ended devices with a resonance order of 100 to 200, depending on design. That is to say, the light wavelength fits around the ring circumference 100 to 200 times. The Q-factor of a typical MRM is in the 3000–7000 range, creating a very narrow optical resonance. Therefore, the MRM must be tuned to a specific resonance condition and held there accurately.

This sensitivity offers a key benefit: each MRM in a bus can act on a separate WDM wavelength with only 0.1 dB noise penalty due to optical crosstalk. Therefore, the MRM is particularly useful for a compact WDM TRx [8].

4.2 Wavelength Accuracy

The relative wavelength detuning between the MRM resonance and the laser affects link engineering metrics such as the OMA, ER, and transmitter-referenced receiver noise penalty (e.g., TDEC or TDECQ as defined in IEEE 802.3 standards). Figure 17 shows the OMA of our vertical-junction MRM design versus detuning. The horizontal axis is normalized to the FSR of the MRM, which is 6.6 nm. The OMA varies by 0.5 dB peak-to-peak over ± 0.003 FSR, which is equivalent to 20 pm or 4 GHz.

Lasers are routinely locked absolutely to such wavelength accuracy in DWDM systems. However, absolute wavelength locking is unnecessary in our application, where we require only that the MRM tracks the laser wavelength drift. We do this by varying the heating of the MRM using a heater (proper thermal design and control mechanisms are required), which shifts the wavelength by ~ 170 pm/mW.

Three factors need to be considered:

- Initial relative wavelength. We address this by redundant MRM buses and (for WDM) cyclic wavelength assignment [8].
- Wavelength drift over life due to environment temperature and aging. The MRM heater can compensate for about 1.0 nm drift (0.15 FSR) between the MRM and laser wavelengths.
- Pattern-dependent self-heating [20]. This is addressed by open-loop pattern-dependent heater driving.

We further discuss these factors later in this paper.

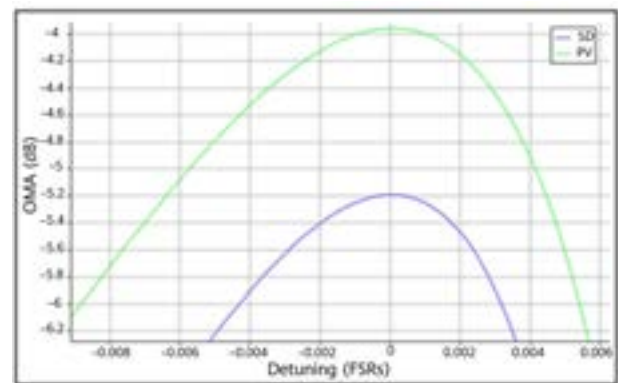


Figure 17 OMA of MRM versus wavelength detuning, for vertical junction design

4.3 WDM Bus Architecture

MRM can be used in a parallel architecture for gray-wavelength fiber ribbon applications and CWDM applications (DR-type and FR-type, respectively, using IEEE 802.3 terminology).

For DWDM, an MRM bus is advantageous and a unique capability of wavelength-resonant modulators. Figure 18 illustrates such a bus. As shown in the figure, continuous-wave light containing multiple wavelengths enters the bus. Each wavelength interacts with only one MRM, passing by all the other MRMs.

The wavelengths must be accurate to a fraction of an FSR but do not need to be in the same resonance order. The

wavelengths (strictly, optical frequencies) must be in the following grid:

$$f[n] = f_0 + k[n] \cdot \text{FSR} + n \frac{\text{FSR}}{N} \quad (4)$$

Here, $f[n]$ is the optical frequency of the n 'th wavelength, f_0 is an arbitrary offset, $k[n]$ is an arbitrary integer, and N is the total number of wavelengths.

If we assume that there are eight wavelengths, the FSR is 6.6 nm, and the simplest case of $k[n] = k_0$ such that all wavelengths are in the same order, this is simply a 0.8 nm DWDM grid. The radius of each ring increases by about 0.1%, meaning that the nominal wavelengths of the MRMs precisely cover a full FSR. In this case, the modulation crosstalk with our MRM design is less than 0.1 dB between any two wavelengths.

Because the MRMs in this architecture are closely packed spectrally, they consume less power for heating and meet reliability targets more easily.

Due to fabrication variations, the absolute resonance wavelength of MRMs varies from wafer to wafer and die to die. However, since fabrication is locally uniform, MRM buses on a given die have well-defined relative resonance wavelengths. Because the wavelengths can be assigned cyclically, as illustrated in Figure 19a, the bus architecture is immune to the absolute wavelength. From the perspective of an Ethernet stack, this is equivalent to lane re-ordering in the physical coding sub-layer.

4.4 MRM Control

MRM control is achieved by monitoring optical power at low speed on a weakly tapped secondary optical coupler on each ring [21]. This tap only contains the wavelength that is close to resonance with the MRM, as illustrated in Figure 17. Our training algorithm applies a short-run-length pattern, scans the heater of each MRM in turn, and monitors the

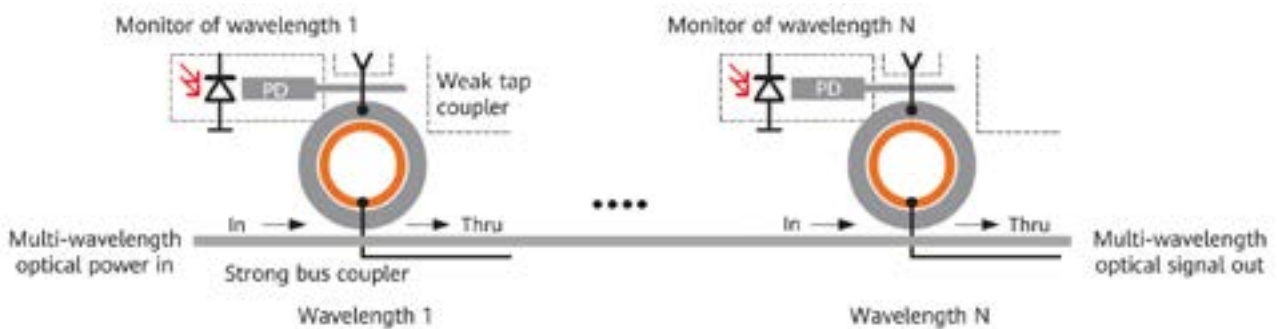


Figure 18 MRM bus for WDM operation

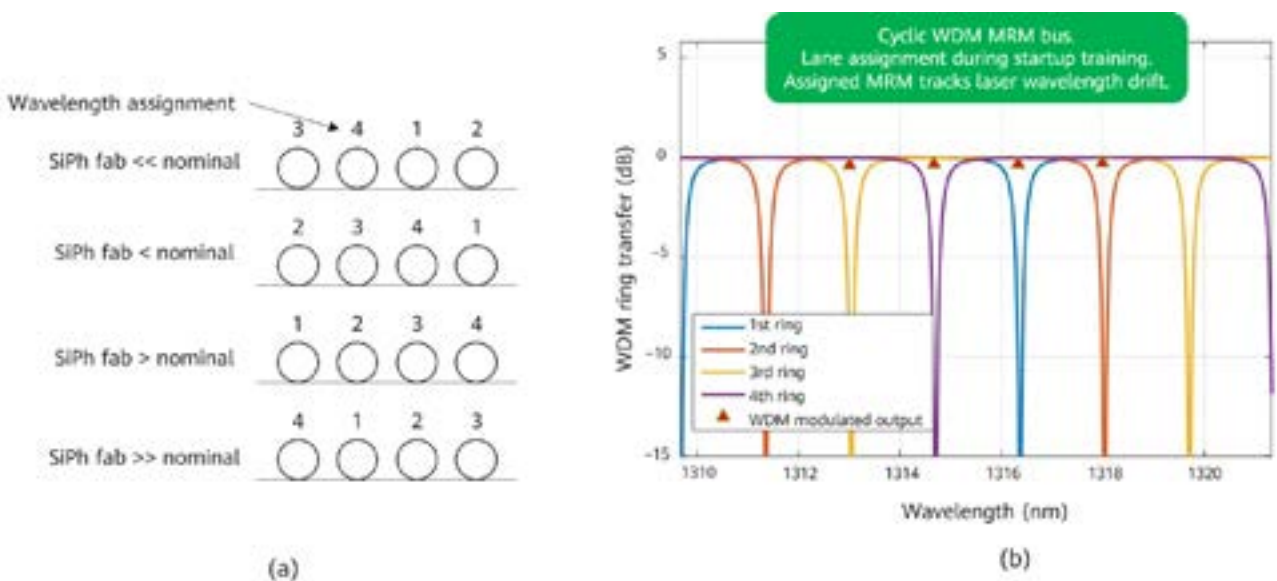


Figure 19 Example operating conditions for a 4-wavelength MRM WDM bus. (a) Cyclic laser wavelength assignment implemented during training sequence – the diagram illustrates the four possible end-states; (b) Operating wavelengths [triangle markers] and MRM transmission spectra [smooth lines] for one end-state. In this scenario, all wavelengths are in the same resonance order $k[n] = k_0$.

DC optical power tapped at each MRM to find the correct operating heater power of the MRM. The algorithm adds allowances for environmental temperature changes and drift over life. In this way, it ensures that there is never an out-of-service retraining event. As such, the worst case MRM temperature is minimized.

Figure 20 illustrates the emulated training sequence for one MRM with four CW input lasers. Training and initial lock is performed with modulation off. The emulation is run with 11.9 dBm input power per laser and +70°C ambient temperature.

For the first 4.3 ms, the controller ramps the heater power down in a search ramp. The resonant wavelength, temperature, and insertion loss follow a sawtooth pattern, because self-heating by light absorbed in the MRM briefly locks the MRM onto each wavelength in turn. This can be seen as the four plateau levels in the wavelength graph. After the search ramp is completed, the controller determines which wavelength to assign to this MRM (1308 nm in this case). From 4.3 ms to 8.2 ms, the controller ramps the heater power down to lock the MRM to the

assigned wavelength. The process may then be repeated for other MRMs on the bus (not shown).

At 8.5 ms, a 56 Gbaud SSPRQ stress test pattern is applied, and the MRM remains locked. Figure 19b shows the final wavelength assignment state after training all four wavelengths onto their assigned MRM.

To maintain the OMA within 0.5 dB of peak, a DAC with ~10 bits ENOB will be sufficient for each lane. This can be easily implemented in the CMOS driver. A control speed of ~5 MHz is sufficient to accommodate training, pattern-dependent self-heating, laser wavelength drift, and environmental changes. We can improve the tracking accuracy of symbol pattern statistics by providing a more accurate time-dependent reference to the control loop [22].

While our architecture monitors insertion loss at low bandwidth, other more sophisticated methods monitor the OMA directly. In principle, such methods provide the best performance [23]. However, an OMA monitor circuit requires a near-line-rate front-end to the monitor circuit, which increases CMOS power consumption and die footprint.

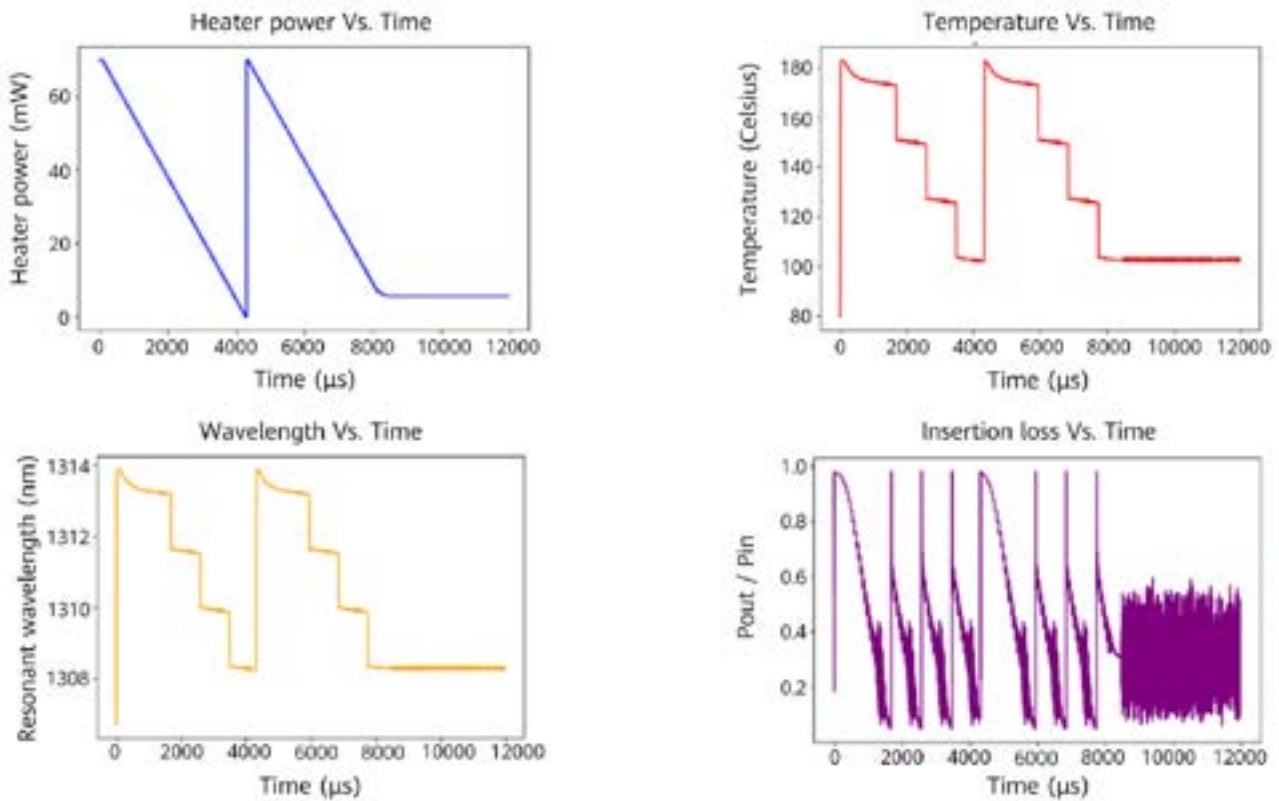


Figure 20 Training an MRM to one of four incident laser wavelengths. Top: Heater power applied to MRM and MRM instantaneous temperature, versus time. Bottom: MRM instantaneous resonant wavelength and insertion loss, same time axis. After 8.5 ms, the MRM is locked, and operates correctly with a stress-testing SSPRQ pattern.

4.5 Temperature Limits and Reliability

The wavelength budget and temperature range increase due to factors such as fabrication uncertainties, uncooled operating environments where the temperature of the optical engine substrate ranges from 0°C to +90°C, and where laser wavelength plan and aging are involved; and packaging and development issues. With a poorly chosen architecture, the total temperature range over operating life would be more than 1 FSR in some configurations, which would be beyond the reliable temperature limits of SiPh.

Tuning an MRM is performed by passing current from a controller through a heating layer in the MRM (the dotted red region of Figure 21a layout). Heater power requirements are similar to an MZM (30 to 60 mW/FSR versus 30 to 60 mW/2 π). This is due to the length-independence scaling law of thermo-optics. However, for the MRM, the temperature change can be large (95°C/FSR for a 10 μ m-radius MRM, and even larger for smaller and faster MRMs). Care must therefore be taken to avoid exceeding the electro-migration temperature limits in the metal wiring. We have adopted several solutions to prevent this failure mode.

First, the MRM is laid out with metals far away from the heater zone. This is facilitated by a skinny racetrack layout, such that heat is kept away from metal traces and vias, as illustrated in Figure 21b. Foundry CMOS rules typically limit the metal operating temperature to +130°C, although the TiN material used for the heater is capable of much higher temperatures.

Second, considering Figure 20, the use of an MRM bus facilitates a low heater power at the maximum environment

temperature. This applies to not only a WDM bus, but also a non-WDM interconnect for which we use redundant MRMs. The heater operating power in Figure 20 is only 7 mW in normal operation after the training is complete, and the MRM runs at only +105°C, which is well below SiPh failure limits.

Note that the heater power consumption at the maximum environment temperature is only 0.07 pJ/bit at 100 Gb/s. This is equivalent to 3.5 W for a 50 Tb/s ASIC.

4.6 Pattern-dependent Self-Heating

Standards-compliant testing of optical interconnects includes test patterns such as SSPRQ that have long consecutive identical digit (CID) groups. This requires careful handling by the MRM heater control loop.

The MRM absorbs more optical power for a low output symbol (e.g., PAM-4 0-level) than for a high output symbol (e.g., PAM-4 3-level). Light is absorbed by dopants and by 2-photon absorption, causing self-heating in the MRM. The optical input power required for a standards-compliant link is ~10 mW. Because the absorbed optical power is comparable to the heater tuning power, absorbed optical power causes the MRM to tune itself. The low-frequency content in a stress test pattern is slow enough that the MRM resonant wavelength drifts if not controlled [20].

In an extreme example, the MRM can show bistable behavior on the long-wavelength side of the MRM resonance, which has positive self-heating feedback. This is illustrated by the discontinuity (vertical jump) in Figure 22.

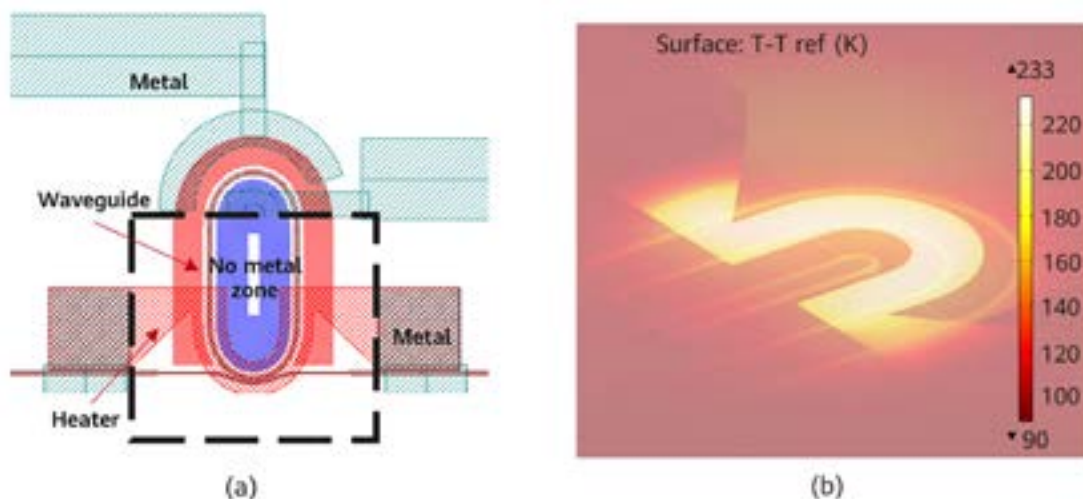


Figure 21 Racetrack MRM to minimize temperature rise in metals. (a) Layout; (b) Temperature model, showing distinct cold zone with metal contacts for RF modulation, and hot zone with heater but without contact metal

We employ several techniques to avoid this behavior. MRM operation is always on the short-wavelength side of the MRM resonance, where the self-heating feedback is negative. Because a low Q factor reduces the self-heating, as the resonance is broader, we set our Q factor as low as possible by optimizing the p-n junction. We apply fast control tracking by damped tuning of the heating power.

For proprietary applications, digital encoding schemes without low frequency content can be optimized to reduce the self-heating by several orders of magnitude. However, doing so comes at the cost of marginal rate reduction and microsecond latency. Our preferred technique is to use constant composition distribution matching codes [24].

In order to boost the modulated optical power to standards-compliant levels, lower laser power may be used by integrating semiconductor optical amplifiers after the MRM output.

As seen from Figure 20, the control mechanism correctly maintains the MRM operating point, even under the most stressful SSPRQ standard test pattern.

5 Conclusion

Today, an MRM with the characteristics necessary for field deployment is a possibility. In this paper, we have demonstrated the design, fabrication, and test of an MRM capable of 56 Gbaud PAM-4 signaling. We have produced devices with a bandwidth greater than 40 GHz using a 200 mm foundry process, and have simulated and proved the control strategy for MRM showing that the known behaviors of self-heating and

initial wavelength selection have practical and robust control solutions.

When using the latest 300 mm wafer-based silicon photonics foundries, mass production of MRM is possible — the fabrication tolerance supports this.

Our industry needs to undergo more changes before the full transition to MRM takes effect. The system architectures and the lasers involved in those architectures affect the practicality of using MRMs. For example, one of the architectures proposed in standards requires an external laser to energize a parallel single-mode array of fibers, each containing a single wavelength. The tuning range for a resonant modulator in that architecture is too large, which will lead to high-power consumption and early failure due to electromigration. The temperatures required to track the input wavelength are also too high. Our industry is only just starting to address the CPO requirements and will need to undergo more changes before it can effectively reverse the electrical power consumption trends of ASICs.

MRMs are best suited for architectures where WDM buses are used. Such an architecture not only provides the highest density for the system (multiple-Tb/cm beachfront), but also simplifies the control aspects of the MRM. We can use a mixture of techniques to best assign a modulator to a laser at startup. In addition to minimizing the power consumption of the system, this approach also prevents MRM device failures.

Manufacturers with integrated laser capabilities have a distinct technology advantage. Through co-integration, it is possible for the MRM to efficiently track the laser drift through the control circuit.

The evolution of high-performance ASICs for artificial intelligence, telecommunication, and computer systems is dependent on the photonic industry's ability to develop innovative low-power and high-density optical interconnects. The long-term viability of CPO depends on our industry increasing the density and decreasing the power consumption of optical modulators.

In the coming years, we expect to witness a transition from the MZM structures commonly used today to the harder-to-control but much smaller and more sensitive MRM structures of tomorrow. In this strategy, we are following a well-traveled path already taken by high-speed interfaces in other domains (e.g., VSR, DSL, PON, Metro, Long Haul, wireless, and Wi-Fi), where control complexity and DSP brought low cost and maximized throughput generation after generation.

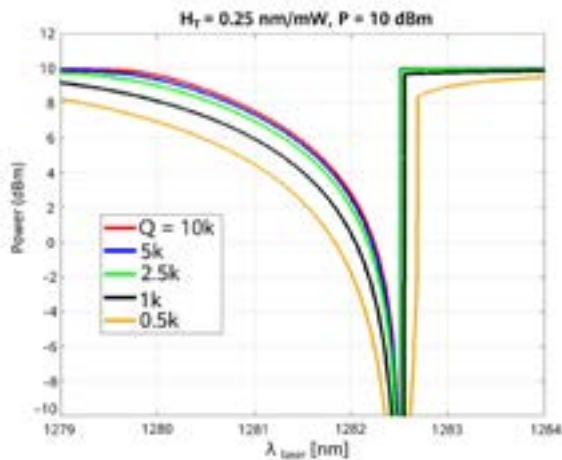


Figure 22 Simulated self-heating of an MRM due to absorbed optical power. Graph shows output optical power versus wavelength for an MRM with 10 mW incident optical power and optical modulation off. The laser wavelength is swept from short-wavelength to long-wavelength. Lines indicate different Q factor. The vertical line on the long wavelength side indicates a bistable jump. We mitigate this, as described in the text. The self-heating is weaker at lower Q factors.

References

- [1] Optical Internetworking Forum (2022). Co-Packaging Framework Document. OIF.
- [2] Chopra, R. (2022). Looking Beyond 400G. IEEE Ethernet 802.3 Working Group, IEEE.
- [3] IEEE Ethernet 802.3 Working Group (2021). Objectives - IEEE P802.3df Task Force. IEEE.
- [4] Huawei Video Library (2022). CloudEngine Cutting-Edge Technology Lecture 6: 400 GE Optical Module. Huawei Technologies Co., Ltd. <https://e.huawei.com/en/videlist>
- [5] Optical Internetworking Forum (2022). Current OIF Work: Common Electrical Interface – 112G-Very Short Reach. OIF.
- [6] Tracy, N. (2021). Preparing-for-a-post-pluggable-optical-module-world. Gazettabyte.
- [7] London, Y., *et al.* (2020). Performance requirements for terabit-class silicon photonic links based on cascaded microring resonators. *Journal of Light-wave Technology*, 38(13), 3469-3477. IEEE.
- [8] Wade, M., *et al.* (2021). An error-free 1 Tbps WDM optical I/O chiplet and multi-wavelength multi-port laser. In *Optical Fiber Communication Conference* (pp. F3C-6). Optical Society of America.
- [9] Fatholouloumi, S., *et al.* (2020). 1.6 Tbps silicon photonics integrated circuit for co-packaged optical-IO switch applications. In *Optical Fiber Communication Conference* (pp. T3H-1). Optical Society of America.
- [10] Norberg, E. (2022). Silicon Photonics with Integrated Laser - Enabling the Paradigm Shift from Pluggable Optics to CPO in Datacenters, in *Optical Fiber Communication Conference* (workshop S1A). Optical Society of America.
- [11] Goodwill, D.J. (2021). The limits of fiber attach to silicon photonics, in workshop "Co-Packaging: How many fibers is too many?", ECOC 2021. Available on request from the authors of the current paper.
- [12] Hu, J., *et al.* (2021). Folded thin-film lithium niobate modulator based on a poled Mach-Zehnder interferometer structure. *Optics Letters*, 46(12), 2940-2943.
- [13] Jin, M., *et al.* (2021). Efficient electro-optical modulation on thin-film lithium niobate. *Optics Letters*, 46(8), 1884-1887.
- [14] Zhang, Y., *et al.* (2020). 200 Gbit/s Optical PAM4 Modulation Based on Silicon Microring Modulator. *European Conference on Optical Communications (ECOC) 2020*.
- [15] Sakib, M., *et al.* (2021). A high-speed micro-ring modulator for next generation energy-efficient optical networks beyond 100 Gbaud. *CLEO 2021*, paper SF1C.3.
- [16] Nedeljkovic, M., *et al.* (2011). Free-Carrier Electrorefraction and Electroabsorption Modulation Predictions for Silicon Over the 1–14-um Infrared Wavelength Range. *IEEE Photonics Journal*, 3(6), 1171-1180, IEEE.
- [17] Yong, Z., *et al.* (2017). U-shaped PN junctions for efficient silicon Mach-Zehnder and microring modulators in the O-band. *Opt. Express* 25, 8425-8439.
- [18] Müller, J., *et al.* (2014). Optical Peaking Enhancement in High-Speed Ring Modulators. *Scientific Reports* 4, 6310.
- [19] Xu, D.-X., *et al.* (2014). Silicon Photonic Integration Platform- Have We Found the Sweet Spot?, *IEEE J. Sel. Top. Quant. Electron.*, 20(4), IEEE.
- [20] de Cea, M., *et al.* (2019). Power handling of silicon microring modulators. *Optics Express*, 27(17), 24274-24285.
- [21] Li, H., *et al.* (2020). A 3D-Integrated Microring-Based 112Gb/s PAM-4 Silicon-Photonic Transmitter with Integrated Nonlinear Equalization and Thermal Control. *International Solid-State Circuits Conference (ISSCC)*.
- [22] Sun, C., *et al.* (2016). A 45 nm CMOS-SOI monolithic photonics platform with bit-statistics-

based resonant microring thermal tuning. IEEE Journal of Solid-State Circuits, 51(4), 893-907.

- [23] Agarwal, S., *et al.* (2016). Wavelength locking of a Si ring modulator using an integrated drop-port OMA monitoring circuit. IEEE Journal of Solid-State Circuits, 51(10), 2328-2344.
- [24] Schulte, P. and Böcherer, G. (2016). Constant Composition Distribution Matching. IEEE Transactions on Information Theory, 62 (1), 430-434.



Technology Research on LNOI-based High-Speed Modulators

Chengcheng Gui ¹, Xiaolu Song ¹, Peijie Zhang ², Tao Gui ², Shuai Yuan ², Keshuang Zheng ², Xuanhao Wang ³, Cheng Zeng ³, Jinsong Xia ³

¹ B&P Laboratory

² Optical Research Dept

³ Wuhan National Research Center for Optoelectronics, Huazhong University of Science and Technology

Abstract

Thin-film lithium niobate (TFLN) modulators deliver outstanding features including low insertion loss, high bandwidth, low driving voltage, high coupling efficiency, and integration capability. They have shown significant application value in technical fields like optical communication. This paper explores the feasibility of empowering high-efficiency coupler, high-speed intensity modulator, and high-speed coherent modulator based on lithium niobate on insulator (LNOI) platform. It is a step toward getting the fabrication and lab verification ready for high-speed modulator with a coupling loss less than 0.6 dB, polarization-dependent loss (PDL) less than 0.1 dB, and 3 dB electro-optic bandwidth greater than 100 GHz.

Keywords

TFLN, LNOI, intensity modulator, coherent modulator, coupler

1 Introduction

As the key device for converting electrical signals into optical signals, electro-optic modulator has exhibited significant value in technical fields such as optical communication, microwave signal processing, optical sensing, and quantum communication [1]. Continuous rise in the data communication capacity leads to higher requirements on electro-optic modulators. High-density integration, low insertion loss, low cost, low energy consumption, and high bandwidth have now become the prerequisites for next-generation electro-optic modulators. Compared with the material platforms used by conventional bulk lithium niobate modulators [2], indium phosphide modulators [3, 4], silicon optical modulators [5, 6], and polymer and plasma modulators [7–9], lithium niobate on insulator (LNOI) stands out with features such as wide optical transparency window, high waveguide refractive index contrast, high electro-optic (Pockels) effect, high bandwidth, and high-efficiency fiber coupling [10–12]. In addition, the fabrication of LNOI devices is in line with its process technology that's capable of quick development and extension. All these make LNOI a competitive material platform for next-generation high-speed electro-optic modulators.

A conventional bulk lithium niobate modulator uses the proton exchange or ion diffusion technology to induce the bulk lithium niobate crystal and generate a small disturbance in the refractive index to form a waveguide. With a metal electrode deployed near the optical waveguide area, the modulator is able to apply an electric field to implement electro-optic modulation. Such a platform has three major drawbacks: (1) A relatively small waveguide refractive index contrast, which is usually 0.02. This reduces the restriction of the optical waveguide over the optical field and thus leads to a relatively large mode area in the optical waveguide. The resulting impact on electro-optic interactions then brings a reduction in the modulation efficiency, significantly increasing the difficulty in reducing the device size and modulation voltage. (2) A high dielectric constant of the substrate material, resulting in a tradeoff between modulation voltage and high bandwidth. This limits the matching between microwave signals and optical signals to some extent, and therefore makes it difficult to implement both low modulation voltage and high bandwidth at the same time. (3) A small refractive index contrast and a large bending radius, restricting high-density integration applications. With the availability of LNOI wafers, LNOI-dedicated nanofabrication technology

has been developed and iterated rapidly. Breakthroughs in LNOI's nanofabrication over the past few years greatly contribute to a mature fabrication process featuring ultra-low loss and high-confinement optical thin-film waveguide [13, 14]. The transmission loss is reported to be less than 0.3 dB/cm [15–17] for thin-film lithium niobate (TFLN) optical waveguides. In addition, the integration of heterogeneous integration and wafer bonding has also recorded a significant development, achieving a waveguide transmission loss of an equivalent level. Among all platforms, some use hybrid waveguides, with optical modes partially limited in TFLN, whereas some use a gradient coupling scheme to implement optical transmission between the TFLN waveguide layer and other material layers [18–21]. The purpose is to limit the optical field as much as possible in TFLN waveguides in order to produce high-efficiency optical modulators. TFLN modulators are made of wavelength-level optical waveguides with a refractive index contrast greater than 0.7, allowing electrodes to be placed several microns close to the edge of waveguides. The optical absorption loss is almost negligible, thereby considerably enhancing electro-optic interaction and the modulation efficiency (VpiL). It is reported that the VpiL of LNOI modulators is in the range of 1.5–3 Vcm [22, 23], which is one magnitude smaller than that of the commercial bulk lithium niobate modulators (15 Vcm) [2]. At the same time, the smaller device size makes it easier to achieve a higher bandwidth through optimized design for lower microwave loss and faster matching. The bandwidth of LNOI-based electro-optic modulation can be easily expanded to more than 100 GHz [24, 25]. To sum up, LNOI modulator is undoubtedly one of the potential candidates for next-generation high-speed electro-optic modulators.

According to different use cases, electro-optic modulators are classified into intensity modulators and coherent modulators. To put it simple, intensity modulators are used to modulate amplitude signals, and are mainly applicable to short-distance optical interconnections. They are used to support low power consumption, high bandwidth, low loss, and high integration. Coherent modulators are used to modulate amplitudes and phases. They are mainly used in long-haul optical transmission and are used to support features such as high bandwidth, low insertion loss, high extinction ratio (ER), and high linearity. In actual applications, efficient coupling between modulators and optical fibers is also an important part of reduced link budget, and has high requirements on the coupling loss and polarization-dependent loss (PDL).

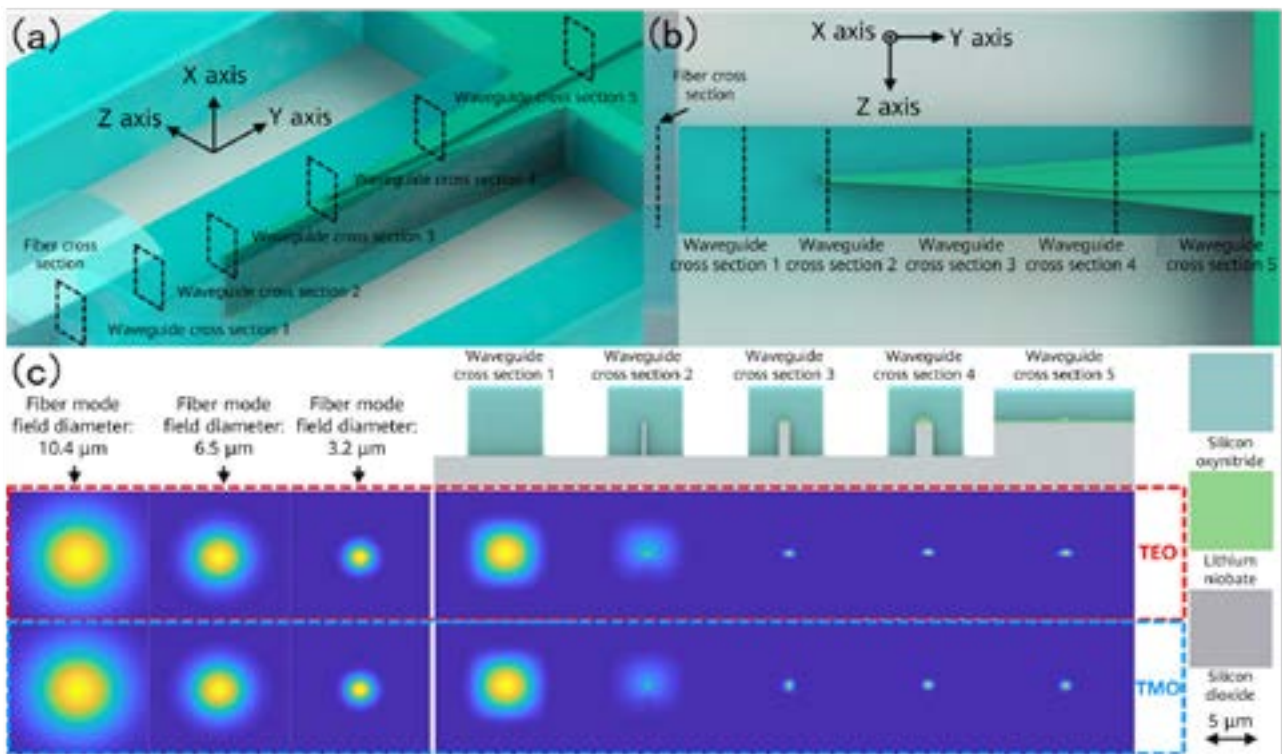


Figure 1 Dual-inverted-cone large mode-size edge coupler

This paper researches the feasibility of LNOI-based high-efficiency edge coupler, high-speed intensity modulator, and high-speed coherent modulator. It explores how to implement a large mode-size edge coupler based on a dual inverted cone structure, so as to achieve coupling loss less than 0.6 dB and PDL less than 0.1 dB. It also investigates the fabrication of LNOI-based high-speed intensity modulators that achieve 3 dB electro-optical bandwidth greater than 100 GHz and on-chip insertion loss less than 2.5 dB, as well as the system verification of 112 Gbps pulse amplitude modulation 4-level (PAM4) and 200 Gbps PAM4 signals. In addition, this paper probes into the design and fabrication of dual-pol coherent modulators, and the experimental verification of single-wavelength 1.6 Tbps high-speed coherent systems.

2 LNOI-based Large Mode-size Edge Coupler

To match the work mode between optical fiber and optical chip device, conventional coupler solutions are divided into two categories: edge coupler and grating coupler. Because grating couplers are reported to have a high coupling loss, design optimization and fabrication are conducted based on edge coupler.

Figure 1 is a schematic diagram for implementing a large mode-size edge coupler based on the dual-inverted-cone structure. It mainly consists of a TFLN dual-inverted-cone structure with large mode size, and a silicon oxynitride waveguide cladding structure. Figure 1a and Figure 1b are a 3D view and a top view showing the structure of the edge coupler, respectively. The silicon oxynitride cladding is used to define the shape of the mode profile and the size of the mode area. The dual-inverted-cone structure can be divided into three parts. First, the TFLN's planar area and the buried layer area are linearly narrowed, whereas the TFLN's ridge waveguide width is kept unchanged. Second, the TFLN's planar area, buried layer area, and TFLN's waveguide area are linearly narrowed to form an upper inverted cone structure, and the widths of the lower TFLN's planar area and buried layer area are decreased linearly to form a lower inverted cone structure. In order to achieve high coupling efficiency, the tip width of the lower inverted cone structure should be small enough. However, due to the limitation of the fabrication process, only the compatible minimum linewidth can be achieved at present. Finally, silicon oxynitride is applied to cover the whole cone area to form the cladding of the dual-inverted-cone structure, so as to define the shape of the mode profile and the size of the mode area. This structure design matches the optical fiber with a mode-field diameter of 6.5 μm . Figure 1c shows the mode change of an edge coupler from the

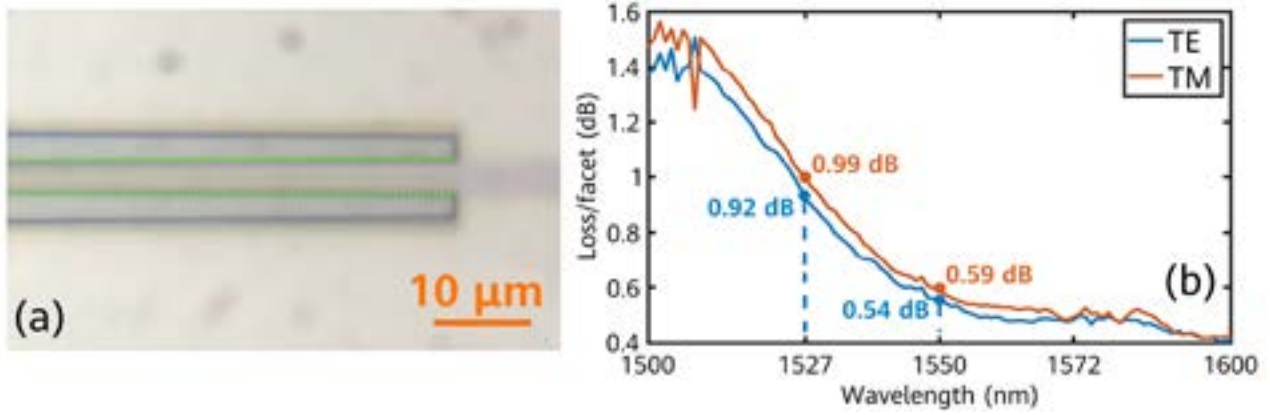


Figure 2 (a) Microscope view of the edge coupler; (b) Coupling loss and wavelength change curves of the edge coupler

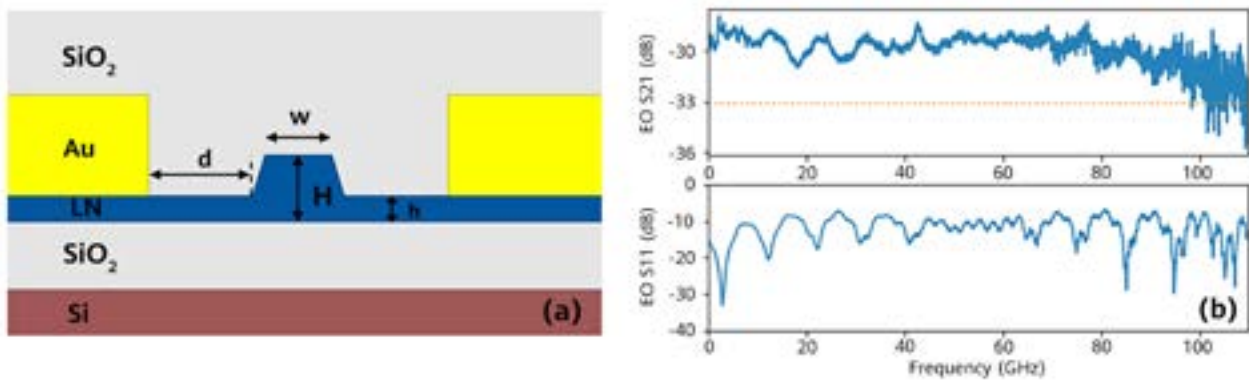


Figure 3 (a) Schematic cross-sectional structure of the intensity modulator; (b) Bandwidth curves of the testing intensity modulator

optical fiber side to the chip side. Waveguide cross section 1 is a schematic structure of the waveguide close to the optical fiber end face. According to the mode field diagram, the mode size is approximately equivalent to the mode-field area (6.5 μm in diameter) of the optical fiber. Waveguide cross section 2 is a tip of a dual-inverted-cone structure, and the optical field diffuses outward at the tip of the waveguide. Waveguide section 3 and waveguide section 4 are gradient areas, and the slave chip mode changes along with the gradient structure. From the change of the mode field from waveguide cross section 5 to waveguide cross section 1, it can be seen that the transverse electric (TE) mode and transverse magnetic (TM) mode diffuse into symmetric spots via a gradient structure. The mode shape is almost the same as that of the optical fiber, thereby achieving high-efficiency coupling.

Figure 2a is a microscope view of the edge coupler, and Figure 2b is the test result of the coupling loss that varies depending on the wavelength. At a wavelength of 1550 nm, the single-end coupling loss is less than 0.6 dB. This value is further reduced as the band extends. Throughout the entire C band, the coupling loss is less than 1 dB and the

PDL is less than 0.1 dB. The overall performance reaches the industry-leading level.

3 LNOI-based Intensity Modulators

Theoretically, the material advantages of TFLN make it easier to achieve low insertion loss and high bandwidth performance for modulators. In order to verify the high bandwidth characteristics, the intensity modulator was first studied. Figure 3a is a schematic cross-sectional structure of the intensity modulator in the modulation area, where an LNOI wafer with a thickness of 500 nm is used. There are five material layers from bottom to top: substrate layer (Si), buried layer (SiO₂), optical guide layer's TFLN (LN), metal electrode layer (Au), and upper cladding layer (SiO₂). For the LNOI modulator in use, the TFLN total thickness H is 500 nm, etching depth is 250 nm, remaining planar waveguide thickness h is 250 nm, device width w is 1200 nm, and spacing d between the waveguide edge and the metal electrode is 3 μm. (Due to TFLN's stable chemical

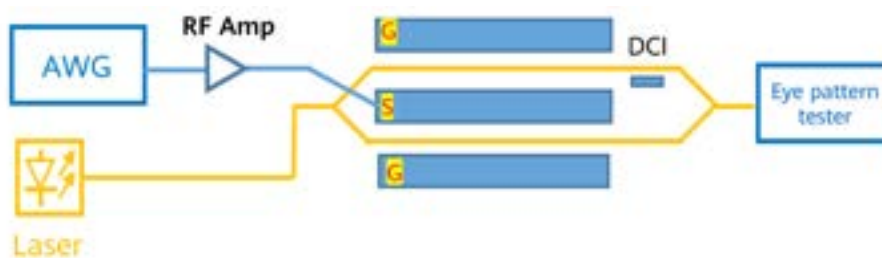


Figure 4 Test environment for intensity modulation

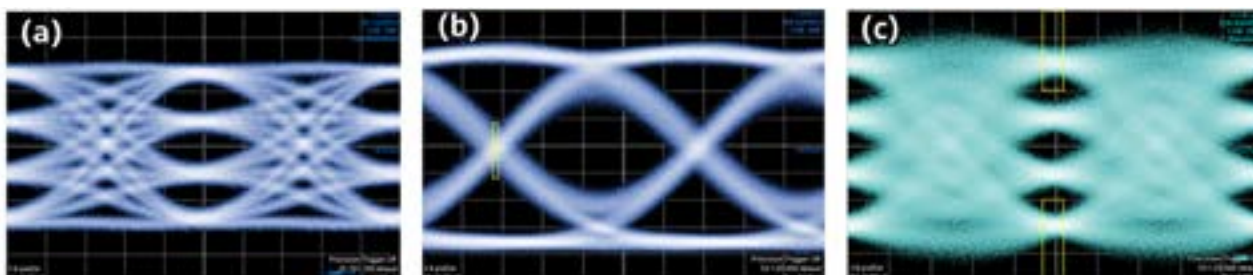


Figure 5 (a) 56 Gbps PAM4 signal pattern; (b) 56 Gbps NRZ signal pattern; (c) 112 Gbps PAM4 signal pattern

characteristics, waveguide etching is quite difficult. This is why the waveguide structure along the etching stroke is trapezoidal). Figure 3b shows the electro-optic bandwidth curves of the testing intensity modulator. According to the figure, the 3 dB electro-optic bandwidth is greater than 100 GHz, modulation length is 9 mm, on-chip insertion loss is less than 2.5 dB, and modulation efficiency is about 2.67 Vcm.

Based on the preceding LNOI intensity modulator, we researched the generation of PAM4 signals at different rates. Figure 4 is the schematic diagram of the experimental setup. An arbitrary waveform generator (AWG) outputs a PAM4 electrical signal. After being amplified by an electrical amplifier (RF Amp), the PAM4 electrical signal is connected to the RF input end of the modulator through a probe. The modulator loads the electrical signal to an optical signal and outputs the optical signal to an optical eye pattern tester.

We start the performance verification by using a low-amplitude signal to drive the intensity modulator to generate a PAM4 signal. During the test, the electrical signal output by the AWG is amplified to an amplitude of 1.5 Vpp by the electrical amplifier and then loaded to the modulator. The generated eye pattern is as shown in Figure 5. Figure 5a is the 56 Gbps PAM4 signal pattern. Without applying the eye pattern tester's equalization algorithm, the ER is 6 dB and the transmitter dispersion eye closure quaternary (TDECQ) is 1.46 dB. Figure 5b is the 56 Gbps non-return-to-zero (NRZ) signal pattern with ER greater than 5.23 dB.

Figure 5c is the 112 Gbps PAM4 signal pattern, where ER is 7.18 dB and TDECQ is 2.79 dB when the eye pattern tester's equalization algorithm is applied.

After the preceding test, we further explore generation of high-speed signals at single-ended 2.8 Vpp drive amplitude. The AWG, as the signal source, outputs 100 Gbps on-off keying (OOK) electrical signals to the electrical driver, where the electrical signals are amplified to 2.8 Vpp OOK signals. The signals are then sent to the intensity modulator through a high-speed probe connection. After the continuous single-wavelength optical signals output by the laser are input to the modulator, the modulator converts electrical signals into optical signals. After that, the signals are output to the eye pattern tester. Due to the bandwidth limitation of the testing device, the performance verification was conducted only on the 100 GBaud system. To accurately reflect the modulator performance, we've provided the eye patterns of the electrical signals, as shown in Figure 5. Figure 6a is a schematic diagram of the 100 Gbps OOK signal output by the AWG. Figure 6b is the electrical eye pattern of the signal amplified by the electrical driver. Figure 6c is the equalized eye pattern output by the eye pattern tester using the optical eye pattern generated by the modulator. Figure 6d, 6e, and 6f are the electrical eye pattern of the electrical signal output by the AWG, electrical eye pattern of 200 Gbps PAM4 signal output by the electrical driver, and optical eye pattern output by the eye pattern tester, respectively. As shown in the figure, the electrical eye pattern deteriorates due to the bandwidth and linearity

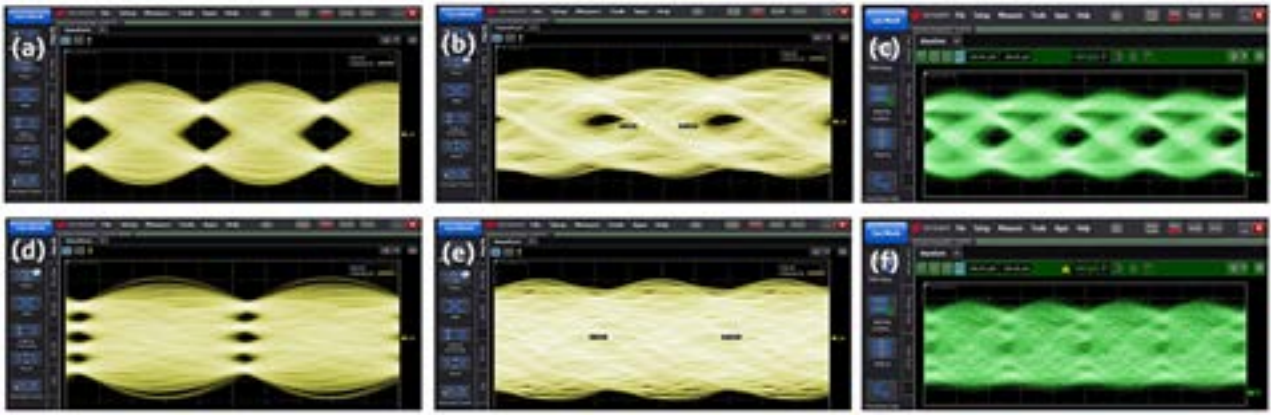


Figure 6 Eye patterns of high-speed signals

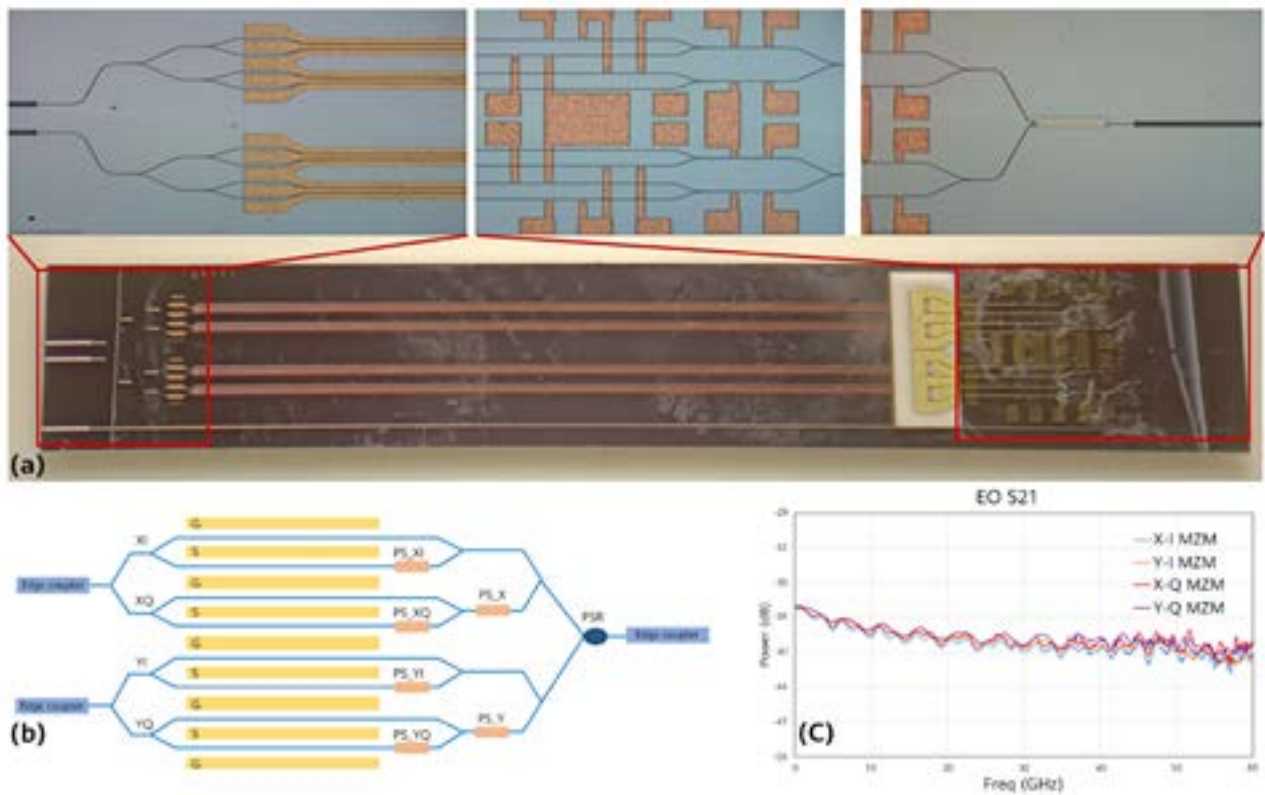


Figure 7 Schematic structure of a coherent modulator

restrictions of commercial electrical drivers. This symptom is more obvious for PAM4 signals. In any case, it is proven that the modulator supports the generation of high-speed signals.

4 LNOI-based Dual-Pol Coherent Modulators

Regarding the next-generation high-speed coherent modulator, we've explored the feasibility of enhancing

system performance using the LNOI technology. Figure 7a is the electron microscope view of a dual-pol LNOI coherent modulator. The two polarization states are marked as X-polarization and Y-polarization, and the four modulators are marked as XI, XQ, YI, and YQ. The modulator consists of three edge couplers, among which two serve as the input of two polarization states, and one serves as the output of the modulator. A total of four Mach-Zehnder (MZ) intensity modulators are used. To reduce the drive amplitude of the modulators, we set the modulation length to 12 cm, on-chip insertion loss to about 12 dB, Vpi to about 2.5 V, and static ER to about 24 dB. Each intensity modulator corresponds to

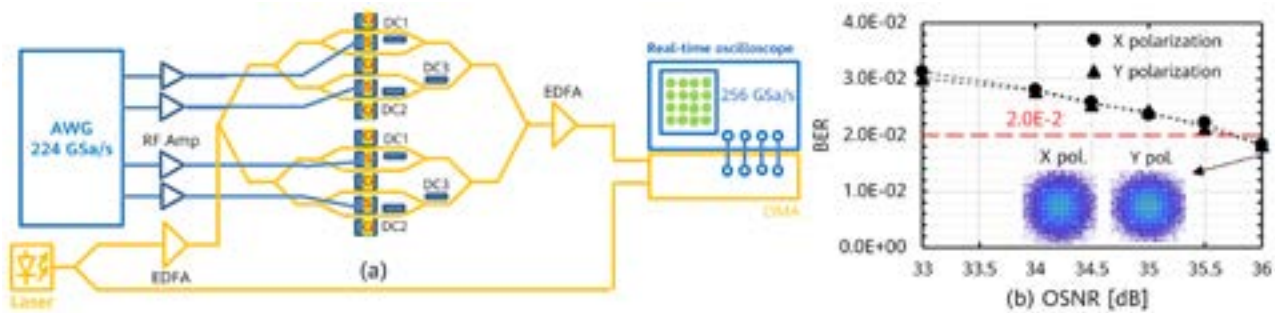


Figure 8 (a) Test environment for the coherent modulator; (b) OSNR and BER curves

a thermal phase shifter, which is used to shift the operating point of the modulator. There is one thermal phase shifter between the I and Q arms of each polarization state. It is used to modulate the 90° phase difference between the two arms. A polarization rotation beam combiner is used to combine the optical signals in the two polarization directions and output the combined signals to the optical fiber through the edge coupler. Figure 7c shows the electro-optic response curves of the XI, XQ, YI, and YQ modulators, where the 6 dB electro-optic bandwidth is greater than 67 GHz (limited by the testing environment without probe loss calibration).

Figure 8a shows the environment for coherent modulator performance testing. An AWG outputs four electrical signals, which are connected to four electrical amplifiers, respectively. The amplified electrical signals are input to four intensity modulators through two GSGSG high-speed probes, and the modulators convert the electrical signals to optical signals. After being transmitted through an optical fiber, the optical signals are input to four balanced phase detectors (PDs), whose 6 dB electro-optical bandwidth is 100 GHz. The PD output end is connected to the real-time oscilloscope, where offline data sampling is performed using a digital signal processor (DSP). After testing the generation and transmission of single-wavelength 130 GBaud 256QAM signals, we tested the optical signal-to-noise ratio (OSNR) and bit error rate (BER) curves, with the forward error correction (FEC) BER threshold set to 2e-2. When OSNR is 36 dB, FEC's BER is lower than 2e-2, indicating that no error occurs after FEC decoding. The test successfully verified zero-error bit rate for 1.6 Tbps transmission over 2 km.

5 Conclusion

This paper is a summary of our exploration and research on LNOI-based high-efficiency couplers, high-speed intensity

modulators, and coherent modulators. We implemented a large mode-size edge coupler based on the dual-inverted-cone structure, achieving a single-end coupling loss less than 1 dB and a PDL less than 0.1 dB throughout the C band. We also investigated the fabrication of LNOI-based high-speed intensity modulators to achieve a 3 dB electro-optic bandwidth greater than 100 GHz and an on-chip insertion loss less than 2.5 dB, and completed the system verification of 112 Gbps PAM4 and 200 Gbps PAM4 signals. In addition, we researched the design and fabrication of dual-pol coherent modulators, and implemented 138 GB 256QAM signal modulation to complete the system verification of single-wavelength 1.6 Tbps high-speed coherent modulators. Through lab verification, the LNOI platform is proven to support implementation of high-speed modulators. Along with the optimization and iteration of the fabrication process, LNOI-based modulators are expected to deliver better system performance.

References

- [1] P. J. Winzer, D. T. Neilson, and A. R. Chraplyvy, "Fiber-optic transmission and networking: the previous 20 and the next 20 years [Invited]," *Opt. Express*, 26, 24190–24239 (2018).
- [2] E. L. Wooten, K. M. Kissa, A. Yi-Yan, E. J. Murphy, D. A. Lafaw, P. F. Hallemeier, D. Maack, D. V. Attanasio, D. J. Fritz, G. J. McBrien, and D. E. Bossi, "A review of lithium niobate modulators for fiberoptic communications systems," *IEEE J. Sel. Top. Quantum Electron*, 6, 69–82 (2000).
- [3] OGISO Y, OZAKI J, UEDA Y, *et al.*, "Over 67 GHz bandwidth and 1.5 V $V\pi$ InP-based optical IQ modulator with n-i-pn heterostructure [J]," *Journal of Lightwave Technology*, 2017, 35(8): 1450-1455.
- [4] Y. Ogiso, J. Ozaki, Y. Ueda, H. Wakita, M. Nagatani, H. Yamazaki, M. Nakamura, T. Kobayashi, S. Kanazawa, Y. Hashizume, H. Tanobe, N. Nunoya, M. Ida, Y. Miyamoto, and M. Ishikawa, "80-GHz bandwidth and 1.5-V $V\pi$ InP-based IQ modulator," *J. Lightwave Technol*, 38, 249–255 (2020).
- [5] M. Li, L. Wang, X. Li, X. Xiao, and S. Yu, "Silicon intensity Mach-Zehnder modulator for single lane 100 Gb/s applications," *Photonics Res*, 6, 109–116 (2018).
- [6] Ding, R. *et al.*, "High speed silicon modulator with slow wave electrodes and fully independent differential drive," *J. Lightwave Technol*, 32, 2240–2247 (2014).
- [7] C. Kieninger, Y. Kutuvantavida, D. L. Elder, S. Wolf, H. Zwickel, M. Blaicher, J. N. Kemal, M. Lauer mann, S. Randel, W. Freude, L. R. Dalton, and C. Koos, "Ultra-high electro-optic activity demonstrated in a silicon-organic hybrid (SOH) modulator," *Optica*, 5, 739–748 (2018).
- [8] S. Ummethala, J. N. Kemal, A. S. Alam, M. Lauer mann, A. Kuzmin, Y. Kutuvantavida, S. H. Nandam, L. Hahn, D. L. Elder, L. R. Dalton, T. Zwick, S. Randel, W. Freude, and C. Koos, "Hybrid electro-optic modulator combining silicon photonic slot waveguides with high-k radio-frequency slot lines," *Optica*, 8, 511–519 (2021).
- [9] C. Koos, J. Leuthold, W. Freude, M. Kohl, L. Dalton, W. Bogaerts, A. L. Giesecke, M. Lauer mann, A. Melikyan, S. Koeber, S. Wolf, C. Weimann, S. Muehlbrandt, K. Koehnle, J. Pfeifle, W. Hartmann, Y. Kutuvantavida, S. Ummethala, R. Palmer, D. Korn, L. Alloatti, P. C. Schindler, D. L. Elder, T. Wahlbrink, and J. Bolten, "Siliconorganic hybrid (SOH) and plasmonic-organic hybrid (POH) integration," *J. Lightwave Technol*, 34, 256–268 (2016).
- [10] C. Wang, M. Zhang, X. Chen, M. Bertrand, A. Shams-Ansari, S. Chandrasekhar, P. Winzer, and M. Loncar, "Integrated lithium niobate electro-optic modulators operating at CMOS-compatible voltages," *Nature*, 562, 101–104 (2018).
- [11] M. He, M. Xu, Y. Ren, J. Jian, Z. Ruan, Y. Xu, S. Gao, S. Sun, X. Wen, L. Zhou, L. Liu, C. Guo, H. Chen, S. Yu, L. Liu, and X. Cai, "High-performance hybrid silicon and lithium niobate Mach-Zehnder modulators for 100 Gbit s⁻¹ and beyond," *Nat. Photonics*, 13, 359–364 (2019).
- [12] A. Boes, B. Corcoran, L. Chang, J. Bowers, and A. Mitchell, "Status and potential of lithium niobate on insulator (LNOI) for photonic integrated circuits," *Laser & Photonics Reviews*, 12(4): 1700256 (2018).
- [13] G. Ulliac, V. Calero, A. Ndao, F. I. Baida, and M.-P. Bernal, "Argon plasma inductively coupled plasma reactive ion etching study for smooth sidewall thin film lithium niobate waveguide application," *Opt. Mater*, 53, 1–5 (2016).
- [14] J. Wang, F. Bo, S. Wan, W. Li, F. Gao, J. Li, G. Zhang, and J. Xu, "High-Q lithium niobate microdisk resonators on a chip for efficient electro-optic modulation," *Opt. Express*, 23, 23072–23078 (2015).
- [15] K. Luke, P. Kharel, C. Reimer, L. He, M. Loncar, and M. Zhang, "Wafer scale low-loss lithium niobate photonic integrated circuits," *Opt. Express*, 28, 24452–24458 (2020).
- [16] R. Wolf, I. Breunig, H. Zappe, and K. Buse, "Scattering-loss reduction of ridge waveguides by sidewall polishing," *Opt. Express*, 26, 19815–19820 (2018).
- [17] R. Wu, J. Zhang, N. Yao, W. Fang, L. Qiao, Z. Chai, J. Lin, and Y. Cheng, "Lithium niobate micro-disk

- resonators of quality factors above 107," *Opt. Lett.*, 43, 4116–4119 (2018).
- [18] P. Rabiei, J. Ma, S. Khan, J. Chiles, and S. Fathpour, "Heterogeneous lithium niobate photonics on silicon substrates," *Opt. Express*, 21, 25573–25581 (2013).
- [19] N. Boynton, H. Cai, M. Gehl, S. Arterburn, C. Dallo, A. Pomerene, A. Starbuck, D. Hood, D. C. Trotter, T. Friedmann, C. T. DeRose, and A. Lentine, "A heterogeneously integrated silicon photonic/lithium niobate travelling wave electro-optic modulator," *Opt. Express*, 28, 1868–1884 (2020).
- [20] P. O. Weigel, M. Savanier, C. T. DeRose, A. T. Pomerene, A. L. Starbuck, A. L. Lentine, V. Stenger, and S. Mookherjea, "Lightwave circuits in lithium niobate through hybrid waveguides with silicon photonics," *Sci. Rep.*, 6, 22301 (2016).
- [21] S. Jin, L. Xu, H. Zhang, and Y. Li, "LiNbO₃ thin-film modulators using silicon nitride surface ridge waveguides," *IEEE Photonics Technol. Lett.*, 28, 736–739 (2016).
- [22] C. Wang, M. Zhang, B. Stern, M. Lipson, and M. Lončar, "Nanophotonic lithium niobate electro-optic modulators," *Opt. Express*, 26, 1547–1555 (2018).
- [23] A. N. R. Ahmed, S. Nelan, S. Shi, P. Yao, A. Mercante, and D. W. Prather, "Subvolt electro-optical modulator on thin-film lithium niobate and silicon nitride hybrid platform," *Opt. Lett.*, 45, 1112–1115 (2020).
- [24] A. J. Mercante, S. Shi, P. Yao, L. Xie, R. M. Weikle, and D. W. Prather, "Thin film lithium niobate electro-optic modulator with terahertz operating bandwidth," *Opt. Express*, 26, 14810–14816 (2018).
- [25] P. O. Weigel, J. Zhao, K. Fang, H. Al-Rubaye, D. Trotter, D. Hood, J. Mudrick, C. Dallo, A. T. Pomerene, A. L. Starbuck, C. T. DeRose, A. L. Lentine, G. Rebeiz, and S. Mookherjea, "Bonded thin film lithium niobate modulator on a silicon photonics platform exceeding 100 GHz 3-dB electrical modulation bandwidth," *Opt. Express*, 26, 23728–23739 (2018).



Antimonide-based High-Sensitivity APDs for 50 Gbit/s Transmission

Junkai Cao, Wei Xiang, Kai Wang, Wei Wei, Shiyong Zhang, Yan Teng
HiSilicon Optoelectronics Co., Ltd.

Abstract

This paper presents an antimonide-based avalanche photodiode (APD) designed and manufactured for high-speed optical communications. With an AlGaAsSb multiplier on the InP substrate, the APD can be grown by using the mass production molecular beam epitaxy (MBE) system and manufactured in the same process as for group III-V materials. Tests show that the APD satisfies the basic requirements of 50G passive optical networks (PONs), with the dark current lower than 5 nA, gain-bandwidth product of 225 GHz, ionization ratio (k) of ~ 0.08 , and sensitivity being -26 dBm at room temperature and -25 dBm at 75°C . In addition, the APD passed the 2000-hour reliability test, making it the first AlGaAsSb optical communications APD for commercial use ever reported in the industry.

Keywords

avalanche photodiode (APD), AlGaAsSb, high-sensitivity receiver

1 Introduction

High-sensitivity APDs are required for high-speed optical communications and laser radar systems. While InP and InAlAs [1–3] are the predominant multipliers used for APDs nowadays, their common disadvantage is that the ionization coefficients (β and α) of holes and electrons are too close to each other, that is, the k value ($k = \beta/\alpha$ or α/β , $k < 1$) is rather high. To achieve better performance, materials with a lower k value are needed. With a lower k value, excess noise can be reduced [4], the avalanche build-up time can be shortened, and the bandwidth of the APD can be extended. Although Si has a relatively low k value, its absorption cut-off wavelength is only 1.1 μm . This is insufficient for optical communications, which require 1.3 μm and 1.55 μm . Using Ge for absorption and Si for multiplication requires a sophisticated heterogeneous epitaxy technique, making the process control rather challenging [5, 6]. Among group III–V materials, Al(Ga)AsSb (lattice-matched to the InP substrate) and both InAlAsSb [7, 8] and InAs [9] (lattice-matched to the GaSb substrate) are of a low k value. However, the bandgap of InAs is too narrow, leading to a high dark current or requiring a low working temperature. Consequently, we compared InAlAsSb and AlGaAsSb and determined that an AlGaAsSb-on-InP scheme is more favorable because it is highly compatible with the current APDs in terms of design, epitaxial growth, and manufacturing technique, with only the InAlAs multiplication region replaced by AlGaAsSb.

Al(Ga)AsSb APDs were first reported by the University of Sheffield [10, 11], and AlAsSb was initially used on the InP substrate. However, in AlAsSb, the composition of group III elements is 100% Al, making it prone to oxidize and eventually leading to a high dark current in the manufactured devices. To improve the dark current performance [12], many subsequent research projects [13, 14] on AlGaAsSb APDs have replaced some Al in AlAsSb with Ga, since the atomic radius of Ga is almost the same as that of Al.

Because the absorption layer for APDs used in high-speed communications must be rather thin, leading to a high tunneling current, and because a large bandgap can reduce the tunneling current, we decided to use $\text{Al}_x\text{Ga}_{1-x}\text{As}_{0.56}\text{Sb}_{0.44}$ (hereinafter referred to as AlGaAsSb) of high Al composition for multiplication.

In this paper, we present our AlGaAsSb APD for 50 Gbit/s transmission. The APD features a low dark current of 5 nA,

k value of ~ 0.08 , gain bandwidth product of 225 GHz, and sensitivity of -26 dBm at room temperature at the data rate of 50 Gbit/s. This is the first AlGaAsSb APD reported in the industry and academia to have passed the reliability tests.

2 Methods

Our APD adopts the separated absorption and multiplication (SAM) structure, with InGaAs as the absorber and AlGaAsSb as the multiplier. It can be grown using the mass production MBE system and manufactured in the same process as for group III–V materials. That is, the epitaxial growth and manufacturing process are almost the same as those for InAlAs APDs.

3 Results

We ran I-V measurements on the APD at room temperature, using photocurrent at 1310 nm wavelength and power of 10 μW . As shown in Figure 1, the APD exhibited breakdown voltage (V_{BR}) of -26 V, and when the voltage is -23 V, the dark current is 5 nA, which is significantly lower than that reported in the literature [13], demonstrating that AlGaAsSb is a suitable material for APDs. Compared with the InAlAs APD that also operates at 50 Gbit/s [15], the dark current level of the AlGaAsSb APD is two orders of magnitude lower. This is mainly because AlGaAsSb has a wider bandgap than InAlAs, thereby reducing the tunneling dark current at the thin multiplication layer. When the voltage is -23 V, the AlGaAsSb APD achieved an avalanche gain of 9.6.

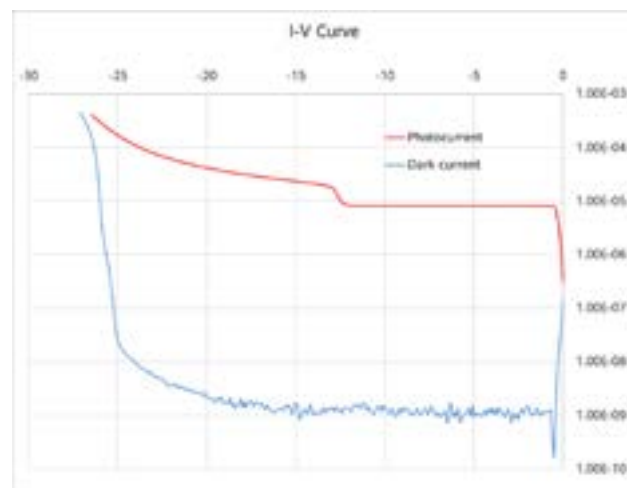


Figure 1 Dark current and photocurrent of the AlGaAsSb APD

Figure 2 shows the gain-bandwidth curve of the AlGaAsSb APD. When gain is 10, the bandwidth is 15.8 GHz; when gain is 30, the bandwidth is 7.5 GHz, resulting in a gain-bandwidth product of ~225 GHz, which is higher than that achieved with our self-developed InAlAs APD.

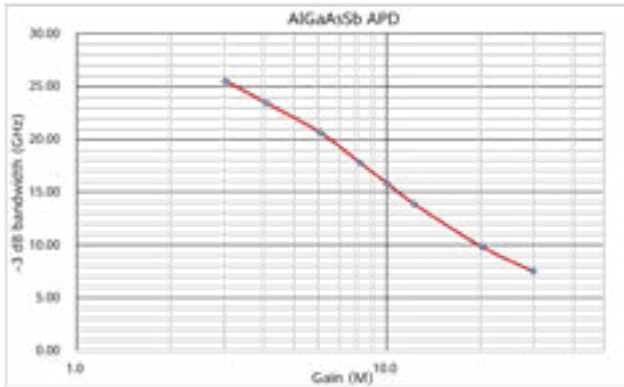


Figure 2 Gain-bandwidth curve of the AlGaAsSb APD

As Figure 3 shows, the temperature coefficient of breakdown voltage (C_{bd}) of the AlGaAsSb APD is 8 mK/V. The C_{bd} value is much higher than that reported in [16], since our APD has a thick InGaAs absorption layer. Meanwhile, the C_{bd} value is significantly less than that of the InAsAs APD [17], with which it has almost the same structure.

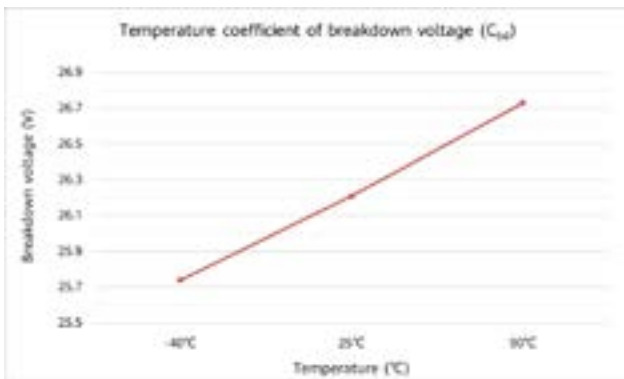


Figure 3 Temperature coefficient of breakdown voltage (C_{bd}) of the AlGaAsSb APD

Figure 4 illustrates the excess noise factor at different gains for our self-developed InAlAs APD and AlGaAsSb APD, respectively. The dotted lines show the measured results and the solid lines are the results calculated by using McIntyre's model [4]. Through fitting, the k value of the InAlAs APD is ~0.22, and that of the AlGaAsSb APD is ~0.08, which is consistent with the results reported in the literature [18].

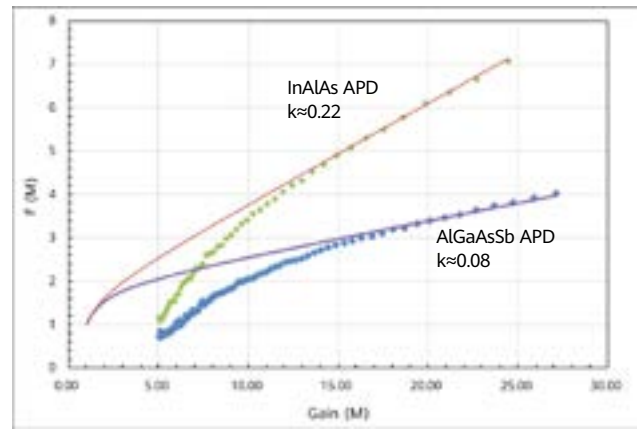


Figure 4 k values of the InAlAs and AlGaAsSb APDs

After the AlGaAsSb APD was packaged, we measured its bit error rate (BER) and sensitivity under the condition of 50-Gbit/s not return to zero (NRZ) transmission. As shown in Figure 5, at the BER of $2E-2$, the APD exhibited sensitivity of -26 dBm at room temperature and -25 dBm at 75°C.

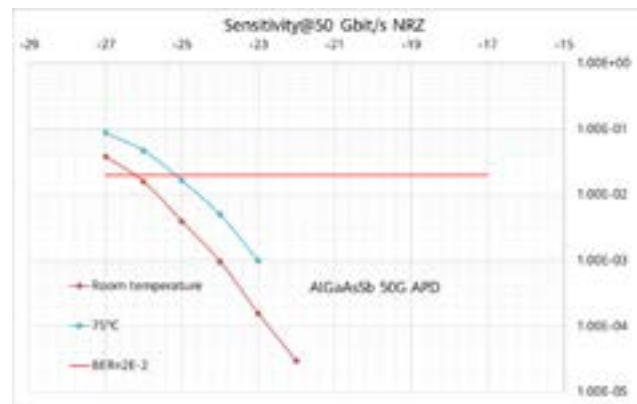
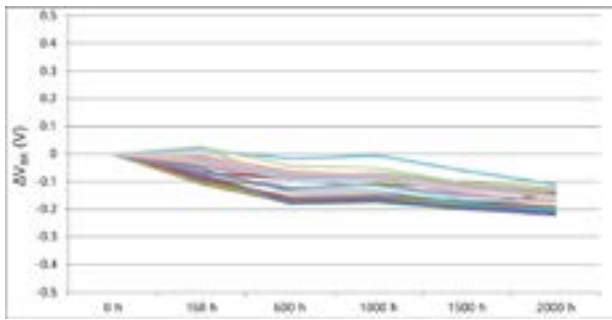
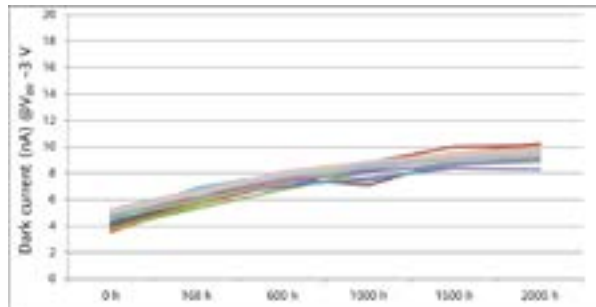


Figure 5 Sensitivity vs. BER for the AlGaAsSb APD

The AlGaAsSb APD also passed all reliability tests defined by common industry standards. Figure 6 shows V_{BR} and the dark current of the AlGaAsSb APD at different aging times. Even after 2,000 hours of aging (175°C, 100 μ A), V_{BR} drops by less than 1 V, and the dark current increases slightly but is still lower than 10 nA. This is the first ever report in the industry and academia to demonstrate that the AlGaAsSb APD can be put into commercial use.



(a)



(b)

Figure 6 (a) V_{BR} and (b) dark current of the AlGaAsSb APD at different aging times

4 Conclusions

This paper presents a 50 Gbit/s AlGaAsSb APD grown using the mass production MBE system and manufactured in the same process as for group III-V materials. The APD demonstrates a k value as low as ~ 0.08 , which is significantly better than InAlAs and InP APDs, and a gain bandwidth product of up to 225 GHz.

Because of its advantageous epitaxy and manufacturing conditions, the AlGaAsSb APD suppresses the high dark current commonly found in APDs with high Al composition. When the voltage is 3 V lower than V_{BR} , the dark current is only 5 nA, far lower than the levels reported by the academia. Moreover, with a large bandgap, the AlGaAsSb APD achieves a lower dark current than InAlAs APDs of a thin multiplication region.

After packaging, the AlGaAsSb APD achieves sensitivity of -26 dBm operating at room temperature and a data rate of 50 Gbit/s. Furthermore, the AlGaAsSb APD has passed the APD reliability tests defined by common industry standards. This is the first time that reliability test results of the AlGaAsSb APD have been reported, demonstrating that the AlGaAsSb APD can be put into commercial use.

Looking ahead, a uni-traveling-carrier absorption region design can be adopted to further reduce the transit time and improve the bandwidth. In addition, further research can be conducted into the composition of Al in AlGaAsSb.

Due to AlGaAsSb having a low k value, it can also improve the performance of 10 Gbit/s and 25 Gbit/s APDs. And, as the epitaxy and manufacturing conditions for InAlAs and AlGaAsSb APDs are very similar, AlGaAsSb can enhance the performance of lower-rate APDs while only slightly increasing costs. Furthermore, with its low k and low dark current characteristics, the AlGaAsSb APD may also bring benefits for laser radar systems.

References

- [1] Campbell, J. C. *et al.*, "Recent advances in avalanche photodiodes," *IEEE J. Sel. Top. Quantum Electron*, 10, 777–787 (2004).
- [2] N. Susa, H. Nakagome, O. Mikami, H. Ando, and H. Kanbe, "New InGaAs-InP avalanche photo-diode structure for the 1–1.6micrometer wavelength region," *IEEE Transactions Electron Devices*, 16, 864–870 (1980).
- [3] M. Nada, T. Yoshimatsu, Y. Muramoto, H. Yokoyama, and H. Matsuzaki, "Design and performance of high-speed avalanche photodiodes for 100-Gb/s systems and beyond," *J. Lightwave Technol*, 33, 984–990 (2015).
- [4] R. J. McIntyre, "Multiplication noise in uniform avalanche diodes," *IEEE Trans. Electron Devices* 13, 164–168 (1966).
- [5] J. Michel, J. Liu, and C. Kimerling, "High-performance Ge-on-Si photometers," *Nat. Photon*, 4, 527–534 (2010).
- [6] Z. Huang *et al.*, "25 Gbps low-voltage waveguide Si-Ge avalanche photodiode," *Optica*, 8, 793–798 (2016).
- [7] M. E. Woodson *et al.*, "Low noise AlInAsSb avalanche photodiode," *Appl. Phys. Lett.*, 108, 081102 (2016).
- [8] A. Rockwell *et al.*, " $\text{Al}_{0.8}\text{In}_{0.2}\text{As}_{0.23}\text{Sb}_{0.77}$ avalanche photodiodes," *IEEE Photon.Technol. Lett.*, 30, 1048–1051 (2018).
- [9] Marshall *et al.*, "High speed InAs electron avalanche photodiodes overcome the conventional gain-bandwidth product limit," *Optics express*, 19, no. 23 (2011): 23341–23349.
- [10] J. Xie *et al.*, "Excess noise characteristics of thin AlAsSb APDs," *IEEE Transactions on Electron Devices*, 59, 1475–1479 (2012).
- [11] Jingjing Xie, "Characterisation of low noise InGaAs/AlAsSb avalanche photodiodes," Diss, University of Sheffield, 2013.
- [12] Yi, Xin, *et al.*, "Extremely low excess noise and high sensitivity $\text{AlAs}_{0.56}\text{Sb}_{0.44}$ avalanche photodiodes," *Nature Photonics*, 13.10 (2019): 683–686.
- [13] Shiyu Xie *et al.*, "InGaAs/AlGaAsSb avalanche photodiode with high gain-bandwidth product," *Optics express*, 24.21 (2016): 24242–24247.
- [14] S. Lee *et al.*, "Low noise $\text{Al}_{0.85}\text{Ga}_{0.15}\text{As}_{0.56}\text{Sb}_{0.44}$ avalanche photodiodes on InP substrates," *Applied Physics Letters*, 118.8 (2021): 081106.
- [15] Nada, Masahiro, *et al.*, "56-Gbit/s 40-km optical-amplifier-less transmission with NRZ format using high-speed avalanche photodiodes," *Optical Fiber Communication Conference*, Optical Society of America, 2016.
- [16] S. Xie, X. Zhou, S. Zhang, *et al.*, "Temperature dependence of avalanche gain in $\text{Al}_{0.85}\text{Ga}_{0.15}\text{As}_{0.56}\text{Sb}_{0.44}$ APD," *2016 Compound Semiconductor Week (CSW)[Includes 28th International Conference on Indium Phosphide & Related Materials (IPRM) & 43rd International Symposium on Compound Semiconductors (ISCS)*, IEEE, 2016: 1-2.
- [17] J. J. S. Huang, H. S. Chang, J. Yu-Heng, *et al.*, "Temperature dependence study of mesa-type InGaAs/InAlAs avalanche photodiode characteristics," *Advances in OptoElectronics*, 2017.
- [18] Pinel, Lucas LG, *et al.*, "Effects of carrier injection profile on low noise thin $\text{Al}_{0.85}\text{Ga}_{0.15}\text{As}_{0.56}\text{Sb}_{0.44}$ avalanche photodiodes," *Optics express* 26.3 (2018): 3568–3576.



Technical Exploration of Tellurium-based Fiber Amplifiers

Zhipeng Zhao¹, Zhixu Jia², Weiping Qin², Guanshi Qin², Shiyi Cao¹, Shengling Wu¹, Shujie Li¹, Yehui Liu¹, Zhiwu Chang¹, Dejiang Zhang³, Xinhua Xiao¹

¹ B&P Laboratory

² National Key Laboratory of Integrated Optoelectronics, School of Electronic Science and Engineering, Jilin University

³ Optical Architecture & Technology Planning Dept

Abstract

The L-band is a low-loss window second only to the C-band in a fiber communication system. To maximize the low-loss transmission window and single-fiber transmission capacity, the L-band needs to be extended to the Super L-band. The bottleneck to achieving this goal is the fiber amplifier for the Super L-band. Due to the impact of excited state absorption (ESA), the existing erbium-doped fiber amplifiers (EDFAs) cannot achieve a net gain at longer wavelengths of the Super L-band. Huawei's Super L-band fiber amplifiers use multi-component fluorotellurite erbium doped fiber (EDF) as the gain medium. When the pump optical power is 910 mW and the signal optical power is 0 dBm, the gain is greater than 16 dB. This lays a foundation for the engineering application of Super L-band EDFAs.

Keywords

fiber, erbium-doped fiber amplifier

1 Introduction

Continuously increasing the single-fiber transmission capacity is a basic requirement of optical communication networks, and expanding the frequency band that can be used on existing optical fibers is a major research direction. Currently, the main transmission window is in the C-band. In China, the C120 band (1524.31–1572.27 nm) is used. The current optical communication networks are systems where the signal-to-noise ratio (SNR) of optical signals is limited. Fiber loss is the greatest impact on the SNR of optical signals. Therefore, frequency bands with low fiber loss are preferred. Based on the loss in mainstream communication fibers [1, 2], the L-band is a preferred candidate following the C-band. In addition, in order to maximize the low-loss window of communication fibers, the L-band needs to extend to as many longer wavelengths as possible, that is, we need Super L-band fiber amplifiers to support the expansion of the optical communication network to the Super L-band.

Currently, the frequency band supported by commercial erbium doped fiber amplifiers (EDFAs) is extended to the 1620–1622 nm [3, 4] in the L-band. The excited state absorption (ESA, that is, transition from $^4I_{13/2}$ to $^4I_{9/2}$) of Er^{3+} ions is a major obstacle that hampers further expansion to the remaining longer wavelengths of the Super L-band. The longest wavelengths an EDFA can support is determined by the intersection point of the stimulated emission cross-section (transition from $^4I_{13/2}$ to $^4I_{15/2}$) and the ESA cross-section. As shown in Figure 1, compared with those in

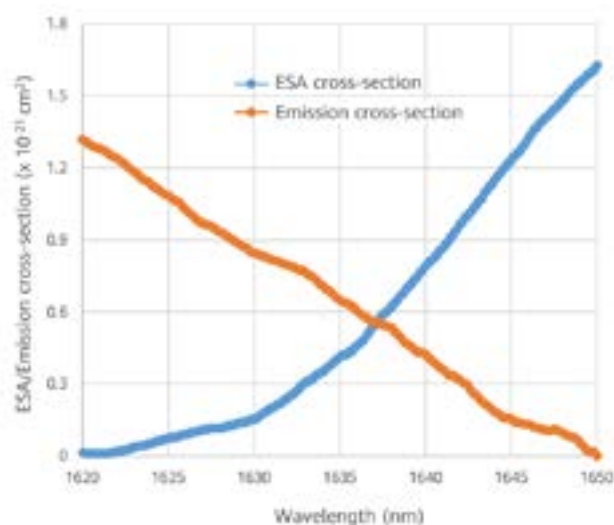


Figure 1 Stimulated emission cross-section and ESA cross-section of Er^{3+} ions in tellurite glass

silica glass, the Er^{3+} ions in the tellurite glass have a larger stimulated emission cross-section and a smaller ESA cross-section in the Super L-band, and the intersection point is located at 1637 nm. This indicates that the longest operating wavelength of tellurite EDFAs can reach 1637 nm, which is 7 nm and 9 nm longer than those of the EDFAs based on fluoride fibers and silica fibers (1630 nm and 1628 nm), respectively [5]. Therefore, erbium doped fiber (EDF) is expected to be used to develop high-gain and wideband fiber amplifiers whose operating wavelength range covers the Super L-band.

In previous reports, wideband optical amplification with a bandwidth of 113 nm (1524–1637 nm) has been implemented in tellurite EDFAs, and the longest operating wavelength of EDFAs has been extended to 1637 nm [6]. This proves the potential of tellurite EDFAs in the development of Super L-band fiber amplifiers. However, the overall gain is low. To further improve the performance of tellurite EDFA, we have optimized the components of fiber host material by adding fluoride and improved the fiber manufacturing process to manufacture fluorotellurite EDF [7–11] with low hydroxyl content, good thermal/chemical stability, and high laser damage threshold. We have also co-doped Ce^{3+} ions to suppress the photon upconversion process. Eventually, the $\text{Er}^{3+}/\text{Ce}^{3+}$ co-doped fluorotellurite glass fiber is used as the gain medium. This is a major technical breakthrough in Super L-band EDFA. When the pump optical power is 910 mW and the signal optical power is 0 dBm, the wideband optical amplification can achieve a net gain greater than 15 dB in the Super L-band.

2 Experiment and Result Analysis

As shown in Figure 2, when Ce^{3+} ions are co-doped, the phonon-assisted energy transfer processes $^4I_{9/2} \rightarrow ^4I_{11/2}(\text{Er}^{3+})$: $^2F_{5/2} \rightarrow ^2F_{7/2}(\text{Ce}^{3+})$ and $^4I_{11/2} \rightarrow ^4I_{13/2}(\text{Er}^{3+})$: $^2F_{5/2} \rightarrow ^2F_{7/2}(\text{Ce}^{3+})$ can implement depopulation of the Er^{3+} ions at excited energy levels $^4I_{9/2}$ and $^4I_{11/2}$, and further increase the population at excited energy level $^4I_{13/2}$. This is conducive to obtaining highly efficient and wideband EDFAs.

Based on the preceding ideas, in the experiment, we first prepared $\text{Er}^{3+}/\text{Ce}^{3+}$ co-doped fluorotellurite glass preform using the sucking method, and then stretched the glass preform to a diameter of approximately 3 mm. Finally, we put the thin rod into an undoped fluorotellurite glass tube to make the optical fiber. Figure 3 is a photo of a prepared $\text{Er}^{3+}/\text{Ce}^{3+}$ co-doped fluorotellurite glass fiber and its cross-

section. The diameter of the fiber core is approximately 7 μm , the outer diameter of the fiber is approximately 122 μm , and the eccentricity is approximately 0.4 μm . Using the cutback method, we measured the fiber loss at the 1980 nm wavelength, which is approximately 0.5 dB/m.

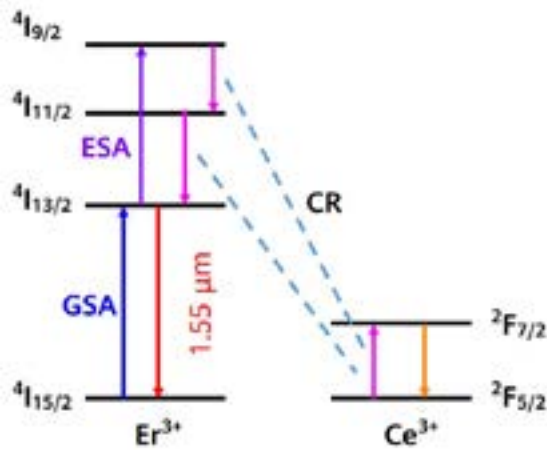


Figure 2 Er^{3+} and Ce^{3+} energy levels

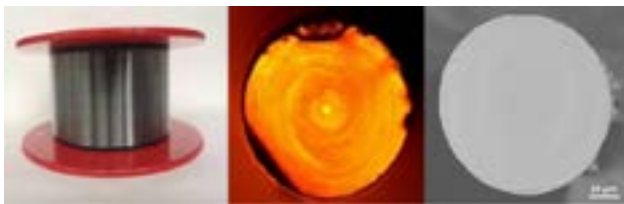


Figure 3 $\text{Er}^{3+}/\text{Ce}^{3+}$ co-doped fluorotellurite glass fiber and its end face

To verify the application potential of the $\text{Er}^{3+}/\text{Ce}^{3+}$ co-doped fluorotellurite glass fiber in the development of the Super L-band fiber amplifiers, a test device shown in Figure 4 is set up to test the amplification performance for the fiber. The signal source is a tunable single-frequency laser with a range of 1470–1660 nm. The signal light and pump light are coupled into single-mode silica fibers through a wavelength division multiplexing (WDM), and then coupled into the $\text{Er}^{3+}/\text{Ce}^{3+}$ co-doped fluorotellurite glass fiber through mechanical splice (the coupling efficiency is approximately 70%). To reduce the impact of reflection on the fiber end face in the optical amplification performance test, the output end of the single-mode silica fiber and the two ends of the gain fiber are cut at a slope angle of 7° – 10° . An optical spectrum analyzer (OSA) is used to record the output spectrum and an optical power meter is used to test the output optical power.

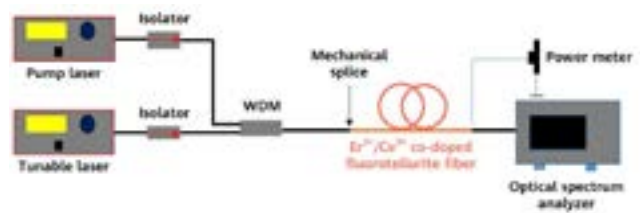


Figure 4 Schematic diagram of an apparatus for testing fiber amplification performance

Figure 5 shows a gain spectrum measured by using different lengths of $\text{Er}^{3+}/\text{Ce}^{3+}$ co-doped fluorotellurite glass fibers as gain media. When the length of the gain fiber is approximately 2.8 m, the measured gain of the Super L-band is greater than 16 dB, the highest gain measured is approximately 22.9 dB, and the corresponding gain ripple is approximately 6.9 dB. The above results show that the $\text{Er}^{3+}/\text{Ce}^{3+}$ co-doped fluorotellurite glass fiber can be used to develop high-gain and wideband fiber amplifiers in the Super L-band.

A Super L-band signal light source is set up, as shown in Figure 6. The amplified spontaneous emission (ASE) source forms 60 channels after passing through an interleaver (ITL). Because the signal power attenuation is severe in this case, an amplifier is added after the ITL to amplify the signals, which are then filtered and shaped by a waveshaper to create a signal source, as shown in Figure 6. A variable optical attenuator (VOA) is added to adjust the signal power to measure different gains of the EDF. An OSA is used to monitor the spectral shape and parameters of the signal source.

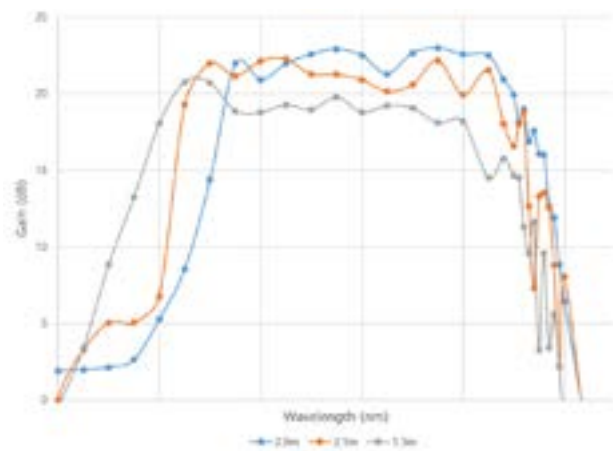


Figure 5 Gain spectrum of $\text{Er}^{3+}/\text{Ce}^{3+}$ co-doped fluorotellurite glass fibers of different lengths

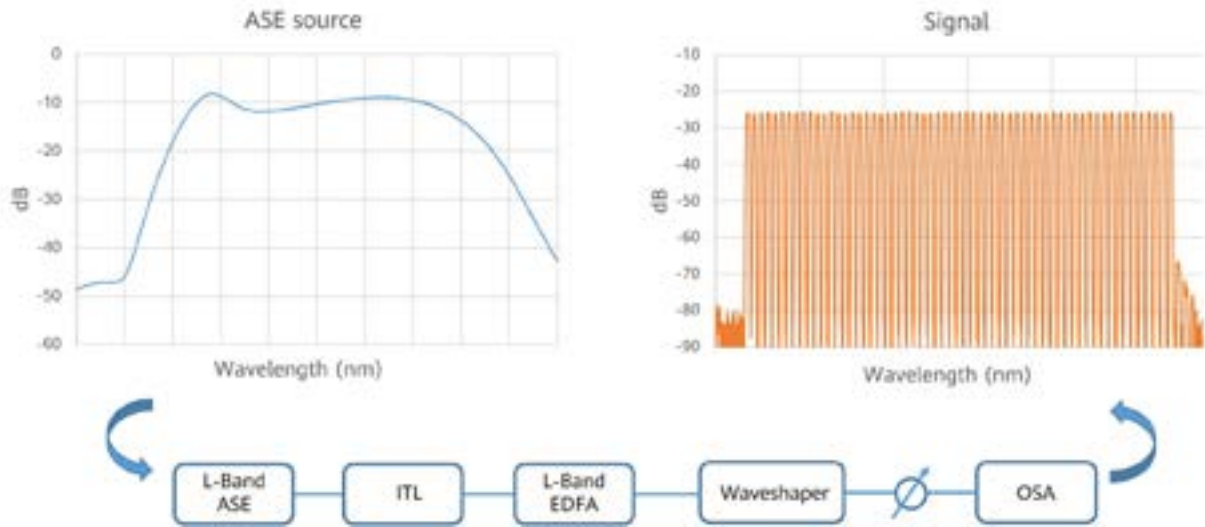


Figure 6 Apparatus diagram of a signal source generation system

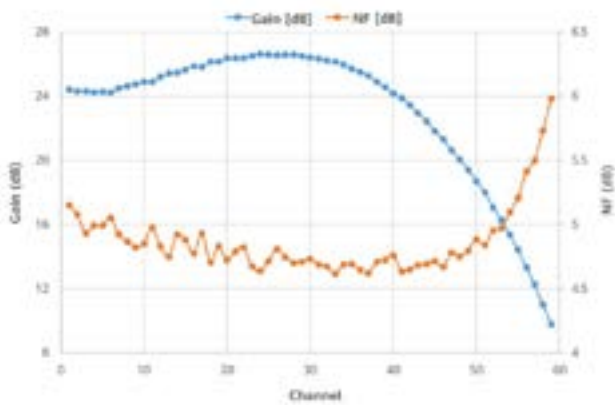


Figure 7 Measured gain spectrum and NF curve of a single-stage bidirectional pump package module

An $\text{Er}^{3+}/\text{Ce}^{3+}$ co-doped fluorotellurite glass fiber package module is also used. The gain fiber in the module is 3.5 m long, and the output fibers at both ends are common single-mode silica fibers. The single-point connection loss of the module is less than 1 dB. Figure 7 shows the curves of the gain spectrum and noise figure (NF) of the single-stage bidirectional pump fiber amplifier based on this module.

When the pump optical power is 800 mW in both directions, the working wavelength range covers the Super L-band and the highest gain measured is approximately 26.6 dB. In the entire Super L-band, the measured NF is less than 6 dB, and the lowest NF is approximately 4.62 dB. By further optimizing the gain fiber and pump laser parameters, and designing and building a cascaded amplification system, the performance of the Super L-band fiber amplifiers can be further improved.

3 Conclusion

By optimizing the components of the fiber, we have proved that the $\text{Er}^{3+}/\text{Ce}^{3+}$ co-doped fluorotellurite glass fiber can be used to develop wideband, high-gain, low-noise Super L-band fiber amplifiers, which are core components for supporting the next-generation large-capacity optical fiber transmission system.

However, due to the impact of ESA, the performance of the Super L-band EDFA deteriorates rapidly at longer wavelengths. Therefore, reducing the ESA in EDFs should be the future research direction and focus of Super L-band EDFA optimization.

References

- [1] <https://www.yofc.com/list/56.html>
- [2] <https://www.corning.com/optical-communications/cn/zh/home/products/fiber/optical-fiber-products.html>
- [3] Tadashi Sakamoto *et al.*, "Wideband rare-earth-doped fiber amplification technologies—gain bandwidth expansion in the C and L bands," NTT Technical Review.
- [4] I.P. Byriel *et al.*, "Silica based erbium doped fiber extending the L-band to 1620+ nm," ECOC2001.
- [5] Yasutake Ohishi, Atsushi Mori, Makoto Yamada, Hirotaka Ono, Yoshiki Nishida, and Kiyoshi Oikawa. "Gain characteristics of tellurite-based erbium-doped fiber amplifiers for 1.5- μ m broadband amplification," Optics letters, 23(4): 274-276 (1998).
- [6] Zhixu Jia, Hang Li, Xiangwei Meng, Lai Liu, Guanshi Qin, and Weiping Qin. "Broadband amplification and highly efficient lasing in erbium-doped tellurite microstructured fibers," Optics Letters, 38(7): 1049-1051 (2013).
- [7] Chuanfei Yao, Chunfeng He, Zhixu Jia, Shunbin Wang, Guanshi Qin, Yasutake Ohishi, and Weiping Qin. "Holmium-doped fluorotellurite microstructured fibers for 2.1 μ m lasing," Optics Letters, 40(20): 4695-4698 (2015).
- [8] Chuanfei Yao, Zhixu Jia, Zhenrui Li, Shijie Jia, Zhipeng Zhao, Lei Zhang, Yan Feng, Guanshi Qin, Yasutake Ohishi, and Weiping Qin. "High-power mid-infrared supercontinuum laser source using fluorotellurite fiber," Optica, 5(10): 1264-1270 (2018).
- [9] Zhenrui Li, Zhixu Jia, Chuanfei Yao, Zhipeng Zhao, Nan Li, Minglie Hu, Yasutake Ohishi, Weiping Qin, and Guanshi Qin. "22.7 W mid-infrared supercontinuum generation in fluorotellurite fibers," Optics Letters, 45(7): 1882-1885 (2020).
- [10] Xiaohui Guo, Zhixu Jia, Yadong Jiao, Zhenrui Li, Chuanfei Yao, Minglie Hu, Yasutake Ohishi, Weiping Qin, and Guanshi Qin. "25.8 W all-fiber mid-infrared supercontinuum light sources based on fluorotellurite fibers," IEEE Photonics Technology Letters, 34(7): 367-370 (2022).
- [11] Yadong Jiao, Zhixu Jia, Xiaohui Guo, Zhipeng Zhao, Yasutake Ohishi, Weiping Qin, and Guanshi Qin. "Third-order cascaded Raman shift in all-solid fluorotellurite fiber pumped at 1550 nm," Optics Letters, 47(3): 690-693 (2022).



Application of Zipper Code in Optical Fiber Communications

Qinhui Huang¹, Xiaoling Yang¹, Huixiao Ma¹, Kechao Huang²

¹ B&P Laboratory

² Optical Communication Technology Lab

Abstract

Zipper code is a promising forward error correction (FEC) code candidate in next-generation optical transmission systems. In principle, zipper code offers a unified framework for spatially coupled codes. Compared with the widely adopted and standardized staircase code, zipper code is able to achieve a better latency-complexity-performance tradeoff by exploiting a higher degree of design freedom.

This article provides a comprehensive investigation on zipper code in terms of code design, decoding algorithms, and implementation architecture. A good compatibility and robustness to the addressed power-efficient techniques make zipper codes stand out in a wide range of scenarios in the "beyond 400G" era, ranging from data center network (DCN) to long-haul transmission.

Keywords

zipper code, FEC, optical fiber communications, beyond 400G

1 Introduction

1.1 Overview of FEC

As a key technology for improving data transmission reliability, forward error correction (FEC) coding has been widely applied to various communication systems. By adding redundancy to data at the transmitter, it ensures that the receiver can rectify the errors that have occurred during data transmission. Based on an ideal assumption, Claude Elwood Shannon, known as the "father of information theory", gave the theoretical performance limit (Shannon limit) of FEC in his milestone article *A Mathematical Theory of Communication* in 1948 [1]. Since then, mathematicians and engineers have started their long journey towards the Shannon limit, resulting in a series of FECs that are practically considered as capacity-approaching, including low-density parity-check (LDPC) codes [2, 3] and turbo codes [4]. The advent of polar codes in 2009 theoretically proves that the Shannon limit can be strictly approached through structured FEC [5].

In real-world applications, FEC design does not blindly pursue the Shannon limit; rather, it is a process of complex system engineering that covers multiple dimensions, including performance, latency, area, and power consumption. After all, strictly achieving the Shannon limit relies on an ideal assumption of infinite code length, which is apparently infeasible in practical applications. Driven by artificial intelligence applications and the continuous promotion of the carbon neutrality strategy in recent years, energy efficiency, in addition to latency, has become increasingly important.

1.2 Optical Fiber Communications and Its FEC Characteristics

In modern communication networks, the application of optical fibers involves both intra- and inter-network connections. According to transmission distance, scenarios can be classified as data center network (DCN), metropolitan area network (MAN), long-haul transmission, and submarine communication. Considering differences in their transmission distance, each scenario is defined with a distinct name, as shown in the following table [6].

Due to the differences in distance, the propagation characteristics and FEC requirements vary depending on scenarios.

Table 1 Scenarios of optical fiber communications

Aspect	DC			MAN	Area	Long-Haul	Submarine
	FR (DCN)	LR (DCN)	ZR (DCI)				
Reach (km)	2	10	80	300	600	4000	10000

In general, to ensure high transmission rates and high stability of optical fiber communications, FEC must have the following features:

- **High decoding throughput:** Currently, the single-wavelength rate of optical fiber communications has reached 400 Gbps, which is 20–40 times higher than that of the high-rate eMBB case (10–20 Gbps) in wireless 5G. Ultra-high throughput poses a severe challenge to the low-power design of decoders.
- **Low error floor:** The re-transmission mechanism, such as HARQ used in wireless communications to improve link reliability, is typically infeasible in optical fiber communication systems because of the high throughput. Therefore, optical fiber communication systems require higher error correction capabilities — in a general sense, the post-FEC bit error rate (BER) is required to be less than 1E-15 (or 1E-13 in Ethernet), which is 10 orders of magnitude lower than that in wireless communications.
- **High coding gain:** In the communications and information theory fields, FEC performance is generally measured by the gap to the Shannon limit (i.e., normalized SNR), and this measurement relies on the post-FEC BER requirement. Considering the same gap, a lower post-FEC BER requirement corresponds to a higher coding gain.
- **Transmission mode:** Unlike wireless communication systems which use burst transmission mode, optical fiber communication systems usually come with continuous transmission mode. Therefore, in order to improve coding gains, optical fiber communication systems typically adopt FECs with a convolutional structure, for example, spatially coupled codes.

1.3 Evolution of FECs in Optical Fiber Communications

Similar to wireless communications, intergenerational differences are obvious in the FEC technology in optical

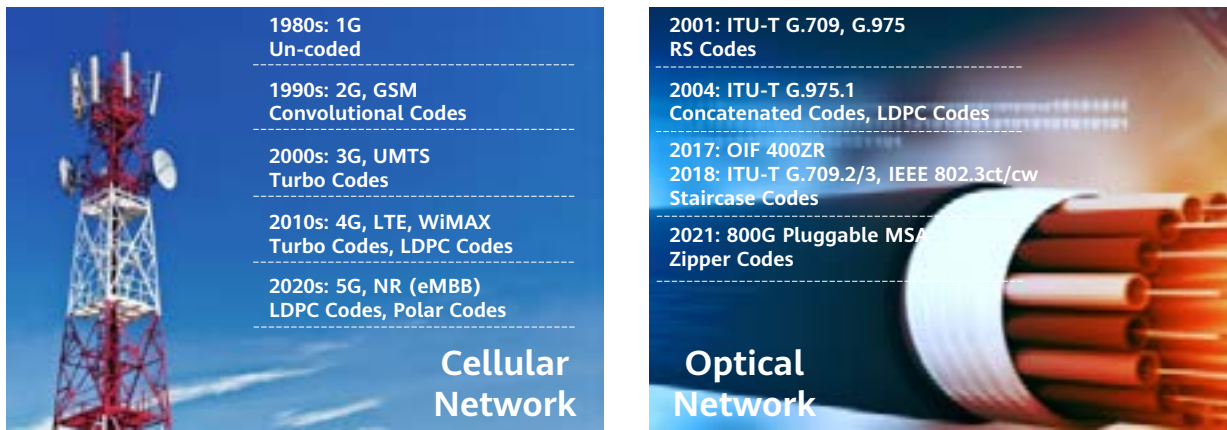


Figure 1 Intergenerational evolution of FEC in cellular and optical fiber communications

fiber communications. In the very beginning, optical fiber transmission used algebraic block codes such as BCH (Bose–Chaudhuri–Hocquenghem) [7, 8] and Reed–Solomon (RS) [9] for error control. As code length increased, the latency and complexity of algebraic codes increased significantly, restricting high performance. Therefore, FECs (such as concatenated codes and block LDPC codes) constructed from a series of simple algebraic codes became the mainstream FECs in the second-generation optical fiber communications. After we entered the coherent communications era in 2010, spatially coupled codes become dominant. Based on concatenated codes and product codes, spatially coupled codes introduce convolutional coupling to meet the high error correction capability while maintaining a limited complexity.

2 Zipper Code: Spatially Coupled FEC in the Beyond 400G Era

Staircase code [10] is one of the most widely used spatially coupled codes in the optical fiber communications field. It uses BCH as component codes and performs staircase coding in row/column interleaving mode. In recent years, staircase codes have been adopted by a number of international standards organizations, such as IEEE 802.3, OIF, and ITU-T, as the standard FEC in 100G/400G systems for different transmission distances. However, as the single-wavelength rate approaches 800G/1.6T, the limitations of staircase codes become prominent due to their fixed coupling structure and limited design space, especially in scenarios involving higher bit rate and lower latency.

2.1 Unified Spatially Coupled Codes

Frank R. Kschischang's team invented *zipper code* in 2019, at University of Toronto [11]. Compared with staircase codes, zipper codes can achieve a better latency-complexity-performance tradeoff by exploiting more flexible coupling, and therefore can be viewed as an upgraded version of staircase codes. As shown in Figure 2a, zipper codes mainly consist of the following elements:

- **Zippering pair:** The zipper code is constructed from a pair of real and virtual buffers, which are located at the two ends of the zipper. The real buffer stores the data frames that actually transmitted, including the raw data of the current frame and the parity data generated after encoding. In contrast, the virtual buffer is filled with the interleaved bits from previous frames of the real buffer. The data in the virtual buffer is not transmitted but will be combined with the raw data of the current frame from the real buffer to generate the parity data.
- **Mapping equation (ϕ):** It is used to define a specific mapping relationship between the virtual data of the current frame and the real data of the previous frame.
- **Component code:** In most cases, algebraic block codes $C(n,k)$, such as Hamming and BCH, are used as component codes, corresponding to the gray arrows in Figure 2. Each component code consists of two parts, one part from the real buffer and the other from the virtual buffer. The two parts are "zipped together" to generate the parity data, which is stored in the real buffer, i.e., the shadow parts in Figure 2.

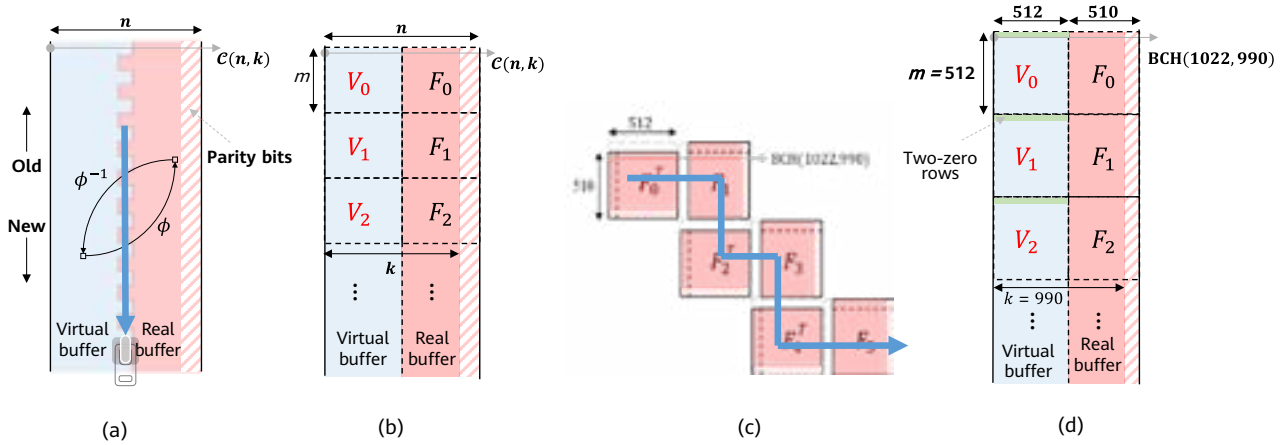


Figure 2 Zipper code structure

Frame structure is also a vital element in this process. In real-world systems, the bit streams of encoded data are generally represented by consecutive FEC frames. As shown in Figure 2b, each real or virtual data frame consists of m half codewords, and the corresponding mapping equation ϕ can be expressed as:

$$V_s = \phi(F_{i_1}, F_{i_2}, \dots, F_{i_D}), i_1, i_2, \dots, i_D \in [S-L, S-1]$$

where S denotes the index of the current frame, and V_s denotes the current frame of virtual buffer; $\{F_{i_1}, F_{i_2}, \dots, F_{i_D}\}$ represents the previous frames of the real buffer from which the data of V_s is mapped, and D represents the coupling depth. Let $C_{range} = 0.5 \times mNL$ denote the coupling range of the zipper code, i.e., the distance between the current frame and its farthest coupled frame.

It can be easily seen from Figure 2c and Figure 2d that, staircase is a special case of the zipper code when the zipper code uses a special frame architecture ($m = n/2$) and a coupling depth of $D = 1$.

2.2 Code Design

In terms of flexibility, staircase codes support only one coupling method, but things are different when it comes to zipper codes. The flexibility of zipper codes allows for lower latency and higher performance.

- Low latency design of zipper codes

Define the coupling degree of zipper codes as $C_{degree} = \frac{N}{2Dm}$, which indicates the average number of bits involved in the encoding of the same group of component codes. From the perspective of code design, zipper codes reduce latency by increasing the coupling degree. Figure 4a and Figure 4b show the cases in which the coupling degree is 1 and 2, respectively. When the coupling degree is doubled, the code coupling range C_{range} is "compressed" to half of the original. In other words, the decoder needs a shorter time to receive codewords and complete one iteration of decoding, leading to a higher convergence speed.

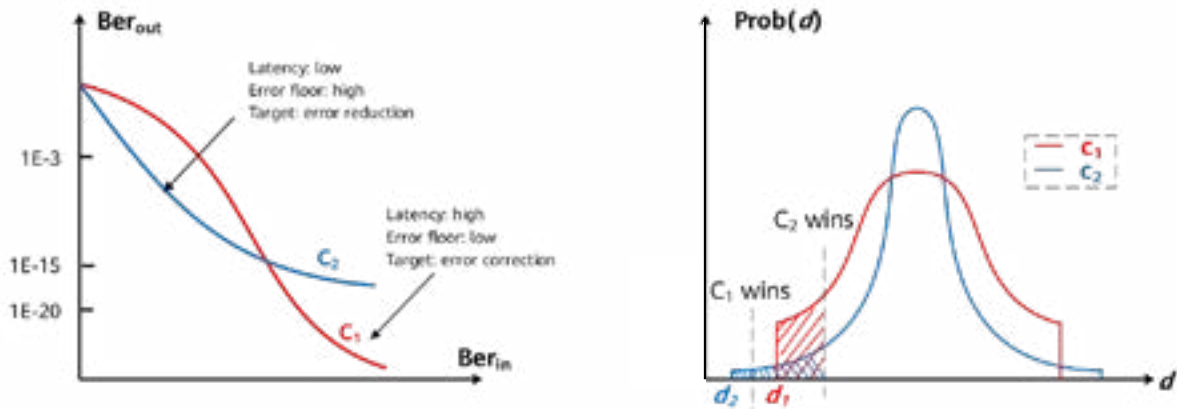


Figure 3 Tradeoff between the convergence speed and the error floor of zipper decoding

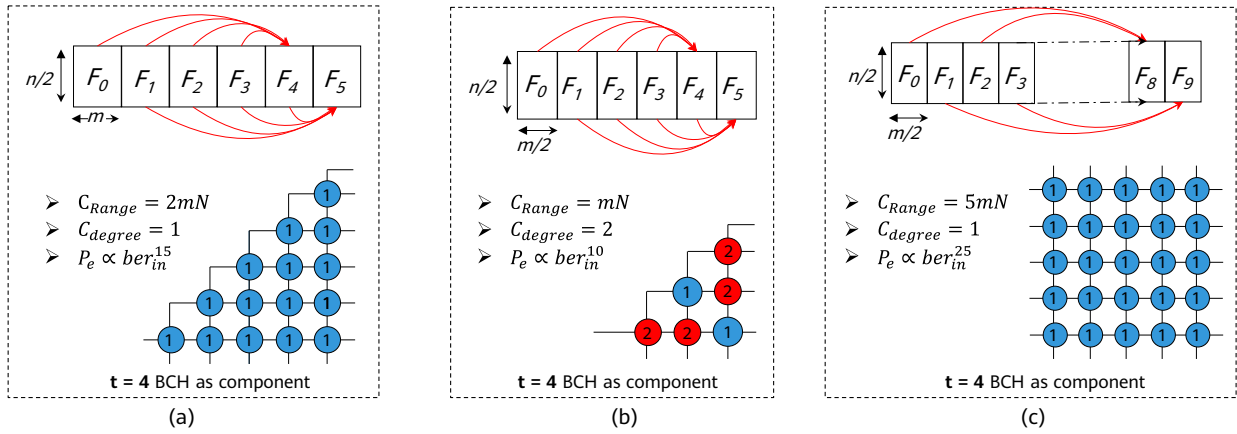


Figure 4 Impact of different zipper code designs on the error floor

However, increased coupling degree usually comes with an "elevated" error floor, as shown in Figure 3. The low-latency zipper codes generally use hard-decision decoding, and their asymptotic performance (error floor) is dominated by the probability of the occurrence of stall patterns. A higher coupling degree means a stronger correlation between component codes, which increases the probability of the occurrence of stall patterns. Figure 4 uses a 4-bit error-correcting BCH code as the component code. When the coupling degree changes from 1 to 2, the number of erroneous bits included in the minimum stall pattern decreases from 15 to 10, and the floor BER also elevates from

$$P_e = A_1 \times (C_m^1)^6 \times (P_{in})^{15} \propto (P_{in})^{15}$$

to

$$P_e = A_2 \times (C_m^1)^4 \times (C_4^3)^2 \times (C_2^1)^2 \times (P_{in})^{10} \propto (P_{in})^{10}$$

where P_{in} is the pre-FEC BER of the zipper code.

- High-performance design of zipper codes
As described above, increasing the coupling degree can reduce latency. On the contrary, by using the bipartite coupling method shown in Figure 4c, the coupling range can be "stretched", so that the asymptotic post-FEC BER introduced by "stalled patterns" will decline from P_{in}^{15} to P_{in}^{25} . This method helps to improve performance in latency-insensitive cases.

High-performance FEC typically employs soft-decision decoding, and the asymptotic performance of soft-decision decoding is determined by the minimum Hamming distance of codewords. [12] proposes an irregular zipper code, which improves the minimum Hamming distance of zipper codes from d_{min}^2 to $d_{min}^2 + d_{min}$ by employing a time-variant mapping equation, where d_{min} is the minimum Hamming distance of component codes.

2.3 Decoding

2.3.1 Decoding Architecture

Similar to convolutional codes, zipper codes are constructed from infinite-length data flows. Therefore, we can capture L consecutive zipper frames as a decoding window, and decode the component codes iteratively in this window to approach the Shannon limit. Similar to the LDPC decoding algorithm, the component codes may be iterated using different scheduling methods, so as to obtain different decoding performance and decoding complexity. The following describes the decoding architectures of three scheduling methods.

- Serial iterative decoding architecture

Serial iterative decoding, also known as *sliding window decoding*, is a common decoding algorithm for zipper codes. Using $L = 6$ and $D = 2$ as an example, six consecutive zipper frames are represented as $F_i, F_{i-1}, F_{i-2}, F_{i-3}, F_{i-4}$, and F_{i-5} , where F_i is the latest frame received, so the corresponding decoding process is: First, decode the component codes from $\{F_i, F_{i-1}, F_{i-2}\}$ and correct the erroneous bits in $\{F_i, F_{i-1}, F_{i-2}\}$. Then repeat the operations for $\{F_{i-1}, F_{i-2}, F_{i-3}\}$, $\{F_{i-2}, F_{i-3}, F_{i-4}\}$, and $\{F_{i-3}, F_{i-4}, F_{i-5}\}$. After all the operations, the decoder outputs data in F_{i-5} . When a new frame F_{i+1} is received, the decoding window encompasses zipper frames $F_{i+1}, F_i, F_{i-1}, F_{i-2}, F_{i-3}$, and F_{i-4} . After iterative decoding of $F_{i+1}, F_i, F_{i-1}, F_{i-2}, F_{i-3}$, and F_{i-4} , data in F_{i-4} is output. The sliding window architecture deals with the decoding iterations serially and can achieve better decoding performance in a small decoding window. Therefore, this algorithm is suitable for low-throughput and low-latency scenarios.

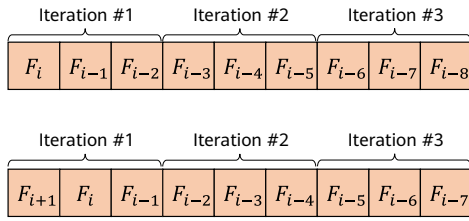


Figure 5 Pipeline decoding architecture

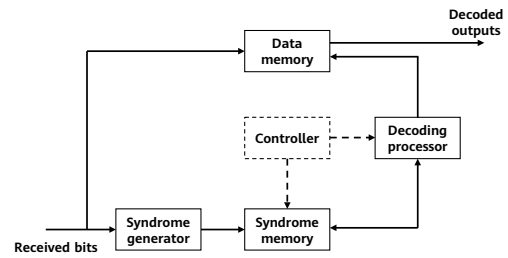


Figure 6 DoD architecture

- Parallel iterative decoding architecture

Parallel iterative decoding, also known as *pipeline decoding*, is an algorithm that helps increase the decoding throughput of zipper codes. As shown in Figure 5, when the zipper frame F_i is received, the first iterative decoding is performed on the component codes from $\{F_i, F_{i-1}, F_{i-2}\}$, the second is performed on the component codes from $\{F_{i-3}, F_{i-4}, F_{i-5}\}$, and the third is performed on the component codes from $\{F_{i-6}, F_{i-7}, F_{i-8}\}$, and then the data in F_{i-8} is output. When the zipper frame F_{i+1} is received, the first iterative decoding is performed on the component codes from $\{F_{i+1}, F_i, F_{i-1}\}$, the second is performed on the component codes from $\{F_{i-2}, F_{i-3}, F_{i-4}\}$, the third is performed on the component codes from $\{F_{i-5}, F_{i-6}, F_{i-7}\}$, and then the data in F_{i-7} is output. This process repeats whenever new frames are received. The component codes in different windows have no data overlapped. This means that during hardware implementation, a large number of parallel decoders may be used to perform parallel decoding on component codes, i.e., high-throughput decoding. The disadvantage is that the decoder resources are configured based on the maximum number of iterations, but in practice the average number of iterations is far less than the maximum number. This leads to low utilization of hardware resources and high power consumption.

- Decoding on demand (DoD) architecture

To meet the power consumption and throughput requirements of high-speed optical networks, [13] proposes DoD, which is a low-complexity high-throughput decoding algorithm aimed at greatly reducing complexity without sacrificing performance. The core of the DoD architecture is to apply a decoding resource pool instead of the maximum resources configured, and dynamically schedule erroneous component codes to the decoder resource pool for decoding. As shown in Figure 6, when a codeword is received, it is stored in the data memory, and after the configured decoding latency, it is output from the

data memory. At the same time, the codeword is sent to the syndrome generator to calculate its syndrome, and the resulting syndrome is stored in the syndrome memory. The controller then schedules the component codes whose syndrome is not zero to the decoding processor. In the event of a successful decoding, the controller updates the corresponding data in the data memory and the corresponding syndrome in the syndrome memory. In this process, the decoding processor serves as a resource pool, which includes N parallel component decoders. For example, when the zipper component code is BCH, the decoding processor may be N parallel BCH bounded-distance decoders. By carefully designing the controller, all component decoders can be kept fully loaded; therefore, a small number of decoders will suffice for performing sufficient iterations on erroneous component codes. This not only meets performance requirements but also significantly saves decoding resources.

To compare the complexity of the two high-throughput decoding architectures, we built a verification platform over Xilinx Virtex UltraScale+ XCVU9P for $t = 4$ BCH component codes with overhead (OH) being 9%. Both architectures are operated at a throughput of 400 Gbps. Figure 7 shows the resource comparison between the two decoders. It can be seen that DoD requires only 1/5 of resources when compared with pipeline decoding.¹

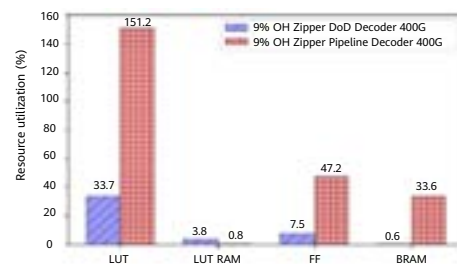


Figure 7 Resource comparison between pipeline decoding and DoD for zipper

¹ Due to the limitations of serial iterative decoding in high-throughput implementation, we only compared the complexity of parallel iterative decoding and DoD.

2.3.2 Decoding Kernel

Sliding window decoding, pipeline decoding, and DoD use different scheduling rules to send component codes to the decoding unit. Based on the method used in the decoding kernel, decoding units are divided into hard-decision decoding, soft-decision decoding, and soft-aided hard-decision decoding.

- **Hard-decision decoding**

Uses bounded-distance decoding (BDD) decoders for iterative decoding of component codes. BDD decoders are easy to implement and can correct $t = \lfloor d_{min} - 1 \rfloor / 2$ erroneous bits, where d_{min} is the minimum Hamming distance of the component code.

By employing BDD decoders iteratively, zipper code may approach the hard-decision Shannon limit. Figure 8a shows the performance of hard-decision zipper codes under different OHs, which are constructed using BCH component codes with different error correction capabilities and different mapping equations ϕ . The gap to the hard-decision Shannon limit is close to 0.6 dB under all settings. This also proves the flexibility of zipper codes — multiple OHs can be flexibly designed and approach capacity at a low-OH region.

- **Soft-decision decoding**

As mentioned above, BDD decoders are able to correct a maximum of $t = \lfloor d_{min} - 1 \rfloor / 2$ erroneous bits. In the case of more than t erroneous bits, BDD decoders may perform *miscorrections*, which will result in error spread and degrading the performance of zipper codes. To improve decoding performance, soft-decision decoding can be applied. Given that the component codes of zipper are mostly algebraic codes like Hamming and

BCH, component decoders can employ ordered statistic decoding (OSD) [14], Chase decoding [15], or their variants for soft-decoding. Combined with turbo-principal-based message passing algorithms, zipper codes are able to reduce the gap to the soft-decision capacity to ~ 1 dB under different OHs, as shown in Figure 8b. The performance can be further improved by introducing irregular component codes and mapping equations ϕ .

- **Soft-aided hard-decision decoding**

Soft-decision decoding is high-performing but results in high complexity and power consumption. There is a lot of research interest on how to improve the decoding performance with limited complexity and low power consumption. Several soft-aided hard-decision decoding algorithms have been proposed in the academia to prevent miscorrections or improve the error correction capability of component codes, so as to improve the performance of zipper codes and achieve a better complexity-performance tradeoff.

For example, [16] proposes an anchor-based algorithm to prevent miscorrections. Specifically, some component codes are marked as anchor codes based on the BDD result, and the decoding results that are inconsistent with anchor codes are discarded. Compared with the traditional BDD, this algorithm improves performance by 0.2 dB–0.3 dB. [17] proposes the iterative BDD with scaled reliability (iBDD-SR) algorithm, which adjusts BDD's hard-decision output to $y_i = L_i + w \times C_i$, where L_i is channel soft information corresponding to the bit and $C_i \in \{-1, 0, 1\}$. If BDD succeeds, $C_i = 1$ or -1 ; if BDD fails, $C_i = 0$. The damping factor w can be optimized separately based on different iterations. [18] proposes soft-aided bit-marking (SABM). This algorithm marks

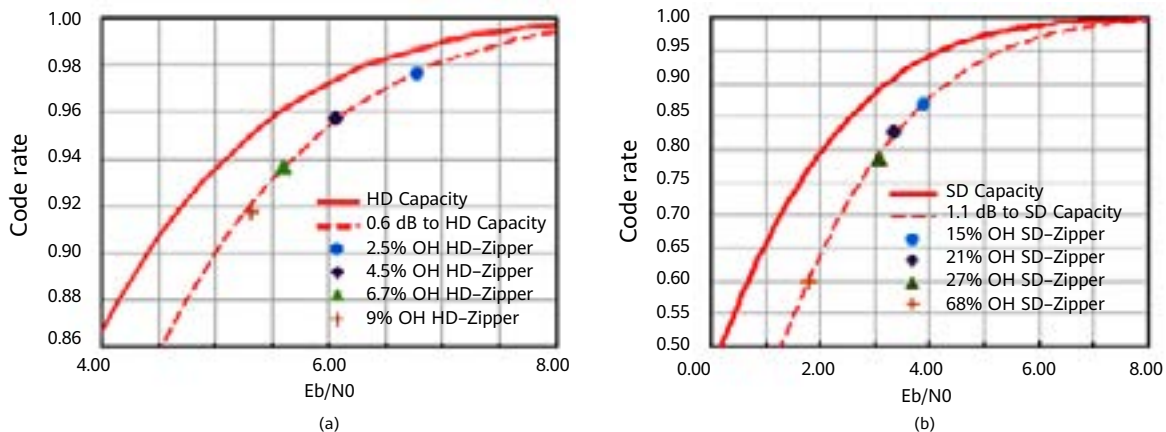


Figure 8 Performance of (a) hard-decision zipper codes (b) soft-decision zipper codes under different OHs

reliable bits and unreliable bits based on the amplitude of soft information received from channels. If a reliable bit is corrected by BDD, the current decoding result is discarded, and an unreliable bit is flipped for another round of BDD. If reliable bit correction occurs again, the decoding result is discarded and decoding of the current component code is terminated. Otherwise, the BDD result is output. By combining the SABM and iBDD-SR algorithms, [19] further improves the performance of zipper codes.

2.4 Extended Applications: Concatenation and MLC

Thanks to its excellent soft/hard-decision decoding performance, zipper codes alone can serve as an FEC scheme in different optical fiber communications scenarios. For example, hard-decision decoding may be applied to DCNs and short-distance metro scenarios that do not demand high performance; soft-decision decoding may be applied to long-distance transmission scenarios such as long-haul and submarine transmission. Thanks to the flexible design, zipper codes can be easily extended to concatenated and multi-level coding (MLC) schemes, and consequently applied to all scenarios of optical communications systems in the beyond 400G era, as shown in Figure 9.

- **Concatenated architecture**

In optical transmission, the time spent on signal

processing is generally controlled within 10% of the link transmission latency. One of the DCN scenarios is short-distance transmission at 500 m–2 km, which poses high requirements on the FEC latency. The current 400GE standard uses KP4 as the FEC scheme. In 2021, IEEE 802.3 positioned 200G/lane as the objective of the next-generation 800G/1.6T standard [20]. As the baud rate of optical components is doubled, link impairment worsens and link budget declines, proving that KP4 alone cannot meet performance requirements. In its white paper released in 2021, the 800G Pluggable MSA group proposed a new scheme wherein HD-Zipper and KP4 are concatenated [21]. As shown in Figure 9a, the host chip side reuses the existing KP4 FEC as the outer code, with the low-latency zipper code concatenated in the optical module as the inner code. On the decoding side, the receiver uses hard-decision decoding of zipper codes to reduce BER to the level supported by KP4, and then uses KP4 to eliminate residual bit errors. This architecture not only meets the requirements of 800GE indicators (latency, performance, power consumption, etc.) but also maintains excellent backward compatibility.

Zipper codes can also be used as outer codes in concatenated systems. In current 400G-ZR standards, the FEC scheme involves concatenating soft-decision Hamming codes with staircase codes [22]. Replacing staircase codes with zipper codes is an efficient way to optimize latency, storage, and complexity, and smoothly evolve DCI scenarios. Relevant details are not described herein.

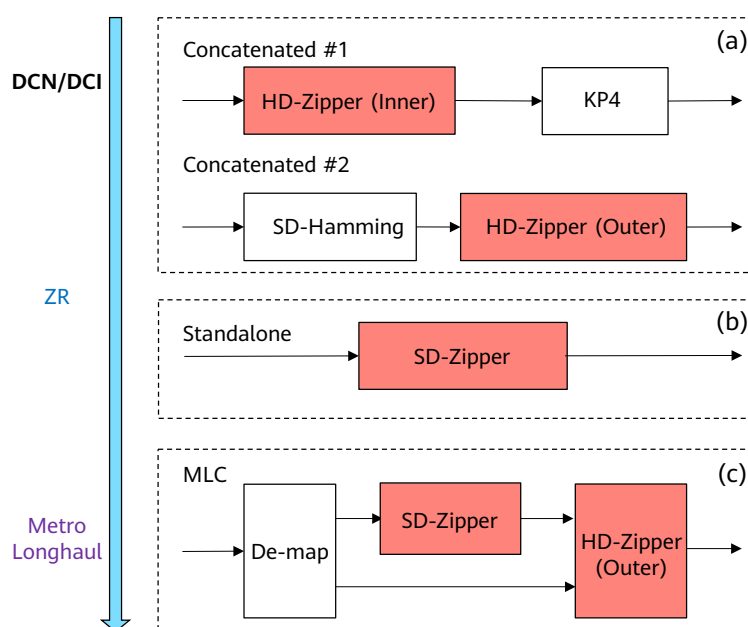


Figure 9 Application of zipper code in optical fiber communications

- **MLC architecture**

Currently, long-haul and metro optical transmissions employ the QPSK and 16QAM modulation schemes. Because of the market trend that single-wavelength rate doubles every 2–3 years, the flex-rate MLC architecture based on higher order modulation (64QAM+) has become a hot research topic in the optical transmission FEC field in recent years. Given the feature that high-order QAM symbols have different error probabilities for each bit, MLC encodes and decodes each layer of bit streams separately. In theory, MLC can achieve high-order QAM constellation-constrained capacity by means of binary coding only. In engineering, MLC can reduce the overall power consumption by lowering the throughput of high-performance soft-decision FEC. Figure 9c is an example of MLC architecture based on zipper codes. In this architecture, the most significant bits have a low BER, and therefore a hard-decision zipper code is used only at the outer layer for protection. As for the least significant bits, a higher-performance soft-decision zipper code is added as the inner code for protection.

The flexibility of zipper codes offers a higher degree of freedom to the design of the MLC architecture, which can be optimized for key MLC features such as rate variability, latency and complexity of soft decoding, and anti-burst capability. [23] proposes a Pareto frontier for designing the MLC scheme based on zipper codes. Compared with bit-interleaved coded modulation (BICM) and other MLC schemes, energy efficiency can be improved at different code rates.

3 Conclusion

This article analyzes the potential of zipper codes in 800G optical fiber communications from the perspectives of code design, decoding algorithms, and implementation architecture. Looking ahead, zipper codes still have some open problems in both theory and implementation, which are worth deeper investigation. In terms of theory, in addition to continuously evolving to higher-order modulation modes, we need to develop a comprehensive distance spectrum analysis methodology to adapt to a broader range of new applications. In terms of implementation, as throughput increases, the implementation cost of zipper codes will shift from logical computing to storage & backend layout, posing greater challenges to in-depth algorithm-architecture-backend design.

References

- [1] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379–423 and 623–656, 1948.
- [2] R. G. Gallager, "Low-density parity-check codes," Cambridge, MA: MIT Press, 1963.
- [3] D. J. C. MacKay and R. M. Neal, "Good codes on very sparse matrices," in *Proc. Cryptography Coding. 5th IMA Conf.*, C. Boyd, Ed., Berlin, Germany, 1995, pp. 100–111, Springer.
- [4] C. Berrou, A. Glavieux, and P. Thitimajshima, "Near Shannon limit error-correcting coding and decoding: Turbo codes," in *Proc. 1993 Int. Conf. Commun.*, Geneva, Switzerland, May 1993, pp. 1064–1070.
- [5] E. Arikan, "Channel polarization: A method for constructing capacity-achieving codes," *2008 IEEE International Symposium on Information Theory*, 2008, pp. 1173–1177.
- [6] Chris Fludger, "Digital signal processing for coherent optical systems," Short Course Notes, OFC 2017.
- [7] A. Hocquenghem, "Codes correcteurs d'erreurs," *Chiffres*, vol. 2, pp. 147–156, 1959.
- [8] R. C. Bose and D. K. Ray-Chaudhuri, "On a class of error-correcting binary group codes," *Inform. Contr.*, vol. 3, pp. 68–79, Mar. 1960.
- [9] S. Reed and G. Solomon, "Polynomial codes over certain finite fields," *J. SIAM*, vol. 8, pp. 300–304, Jun. 1960.
- [10] B. P. Smith, A. Farhood, A. Hunt, F. R. Kschischang, and J. Lodge, "Staircase codes: FEC for 100 Gb/s OTN," in *Journal of Lightwave Technology*, vol. 30, no. 1, pp. 110–117, Jan.1, 2012.
- [11] A. Y. Sukmadji, U. Martínez-Peñas, and F. R. Kschischang, "Zipper codes: Spatially-coupled product-like codes with iterative algebraic decoding," *2019 16th Canadian Workshop on Information Theory (CWIT)*, 2019, pp. 1–6.
- [12] G. Montorsi and S. Benedetto, "Design of

- spatially coupled turbo product codes for optical communications," *2021 11th International Symposium on Topics in Coding (ISTC)*, 2021, pp. 1–5.
- [13] K. Huang *et al.*, "Dynamic decoding of zipper codes," 2021 Optical Fiber Communications Conference and Exhibition (OFC), 2021, pp. 1–3.
- [14] M. P. C. Fossorier and S. Lin, "Soft-decision decoding of linear block codes based on ordered statistics," *IEEE Trans. Inf. Theory*, vol. 41, no. 5, pp. 1379–1396, Sep 1995.
- [15] D. Chase, "A class of algorithms for decoding block codes with channel measurement information," *IEEE Trans. Inf. Theory*, vol. IT-18, no. 1, pp. 170–182, Jan. 1972.
- [16] C. Häger and H. D. Pfister, "Approaching miscorrection-free performance of product codes with anchor decoding," in *IEEE Transactions on Communications*, vol. 66, no. 7, pp. 2797–2808, July 2018.
- [17] A. Sheikh, A. Graell i Amat, and G. Liva, "Iterative bounded distance decoding of product codes with scaled reliability," *2018 European Conference on Optical Communication (ECOC)*, 2018, pp. 1–3.
- [18] Y. Lei *et al.*, "Improved decoding of staircase codes: The soft-aided bit-marking (SABM) algorithm," in *IEEE Transactions on Communications*, vol. 67, no. 12, pp. 8220–8232, Dec. 2019.
- [19] G. Liga, A. Sheikh, and A. Alvarado, "A novel soft-aided bit-marking decoder for product codes," *45th European Conference on Optical Communication (ECOC 2019)*, 2019, pp. 1–4.
- [20] The IEEE P802.3df Project – An Overview (Jan 2022), https://www.ieee802.org/3/df/public/22_01/dambrosia_3df_01_220111.pdf
- [21] 200G PER LANE FOR FUTURE 800G AND 1.6T MODULES, <https://www.800gmsa.com/>
- [22] OIF-400ZR-01.0 – Implementation Agreement 400ZR (March 2020), https://www.oiforum.com/wp-content/uploads/OIF-400ZR-01.0_reduced2.pdf
- [23] M. Barakatain, D. Lentner, G. Böcherer, and F. R. Kschischang, "Performance-complexity tradeoffs of concatenated FEC for higher-order modulation," in *Journal of Lightwave Technology*, vol. 38, no. 11, pp. 2944–2953, 1 June 1, 2020.
- [24] Q. Xie, Z. Luo, S. Xiao, K. Wang, and Z. Yu, "High-throughput zipper encoder for 800G optical communication system," *2021 IEEE International Conference on Integrated Circuits, Technologies and Applications (ICTA)*, 2021, pp. 214–215.



Coded Modulation with Concatenated Polar-Zipper Codes

Ali Farsiabi, Hamid Ebrahimzad, Yoones Hashemi, Bashirreza Karimi, Jin Wang, Chuandong Li, Zhuhong Zhang
Ottawa Optical System Competency Center

Abstract

In this paper, we propose two coded modulation (CM) schemes based on the concatenation of soft-decision polar codes and hard-decision zipper codes: multi-level coding (MLC) and bit-interleaved coded modulation (BICM). In our MLC scheme, we use soft-inner polar codes to protect only a portion of the zipper-encoded bits, thereby significantly reducing the decoding complexity. In our BICM scheme, we propose a new bit-to-symbol mapping and adjust the design of polar codes accordingly. We show that this mapping not only reduces complexity, but also improves the bit error rate (BER) performance. The concatenated polar-zipper codes proposed in this paper form low-complexity forward error correction (FEC) schemes suitable for high-bit-rate optical communication.

Keywords

concatenated codes, polar codes, zipper codes, multi-level coding, bit-interleaved coded modulation

1 Introduction

In today's fiber optic networks, ultra-high-throughput data communication requires reliable forward error correction (FEC) schemes that are capable of low-power and low-latency decoding.

Polar codes [1] have captured considerable attention in the coding community because they can achieve Shannon capacity via low-complexity successive cancellation-based decoding. In particular, the possibility of parallel, low-complexity decoding of certain nodes in the binary tree of the polar codes [2–6] has enabled a tremendous reduction in the decoding latency and power consumption [7].

In order to achieve a bit error rate (BER) less than 10^{-15} with low-complexity and low-power FEC decoding, concatenation of inner soft-decision codes and outer hard-decision codes have gained significant attention [8, 9]. The lower complexity and power consumption of polar codes compared with state-of-the-art FECs such as low density parity check (LDPC) and product-like codes have motivated us to utilize them as the inner code within a concatenated scheme that uses a zipper code as the outer code. In essence, zipper codes [10] are an efficient framework for describing spatially coupled product-like codes. Many of the well-known codes, such as staircase and braided block codes, can be represented with a zipper code structure.

In this paper, we propose two coded modulation (CM) schemes — multi-level coding (MLC) and bit-interleaved coded modulation (BICM) — that utilize concatenated polar-zipper codes to reduce the channel BER to less than 10^{-15} .

We present some preliminaries in Section 2, briefly explaining zipper codes, polar codes, CM, and probing. Then in Section 3, we discuss our proposed CM schemes based on the concatenation of polar-zipper codes. After that, Section 4 evaluates the performance of our proposed schemes based on simulation results. Finally, we offer our closing thoughts in Section 5.

2 Preliminaries

2.1 Outer Zipper Code

In essence, zipper codes [10] are a general framework for describing spatially coupled product-like codes. Many of the

well-known codes, such as staircase and braided block codes, can be represented with a zipper code structure. The three main components of the zipper framework are a zipping pair, an interleaver mapping, and a constituent code. [10] considers a BCH code C^b of length n_b , dimension k_b , and correction capability t_b to be the component code. The zipper buffer is an infinite sequence of BCH codewords $C_0^b, C_1^b, C_2^b, \dots$ that form a semi-infinite matrix with n_b columns. The buffer is divided into a virtual buffer and a real buffer. The bits of the virtual buffer are a direct copy of those in the real buffer through the interleaver mapping ϕ . It is shown that a diagonal interleaver mapping with an equal width of virtual and real buffers in all rows is able to provide very good performance [10]. A zipper code is decoded using a sliding-window decoding algorithm, where M_z consecutive received rows are decoded iteratively, for a maximum iteration of I_{max} using an algebraic BCH decoder [10]. This iterative decoder is invoked whenever a new row is received.

Although diagonal zipper codes can decrease the memory size required for encoding and decoding by about a factor of 2 compared with staircase codes [11], their latency for ultra-low overhead (OH) remains unacceptable for high-throughput applications.

[12] introduces a new interleaver mapping for the zipper framework. According to the new mapping, the required encoding and decoding memory sizes are significantly decreased compared with the original diagonal zipper code. However, such mapping introduces an early error floor caused by certain error locations (known as *stall patterns*) in the real buffer. To tackle this problem, [12] proposes to extend the protection level of each bit in the real buffer from 2 to 3. Consequently, the number of virtual buffers is doubled compared with the single virtual buffer used in the diagonal zipper structure. In addition, a pair of mappings (ϕ_1, ϕ_2) corresponding to the virtual buffers is introduced. By increasing the size of the minimum stall patterns, this pair of mappings brings the error floor-free BER to less than 10^{-15} . The proposed zipper code is called *QDE-zipper code*, where QDE stands for quasi diagonal with extra protection levels.

In this paper, we deploy the QDE-zipper code as the outer code of our proposed schemes.

2.2 Inner Polar Code

A binary polar code $\mathcal{P}(N_p, K_p)$ of length $N_p = 2^{n_p}$ and information length of K_p maps the input vector

$\mathbf{u} = \{u_0, u_1, \dots, u_{N_p-1}\}$ to the output vector $\mathbf{x} = \{x_0, x_1, \dots, x_{N_p-1}\}$ such that $\mathbf{x} = \mathbf{u}\mathbf{F}_2^{\otimes n_p}$, where $\mathbf{F}_2^{\otimes n_p}$ is the n_p th Kronecker power of $\mathbf{F}_2 = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$. The matrix $\mathbf{F}_2^{\otimes n_p}$ synthesizes N_p polarized channels, where the K_p most reliable channels are used to carry information bits. In fact, the vector \mathbf{u} consists of K_p information bits and $N_p - K_p$ fixed bits (referred to as *frozen bits*) that are known to the receiver. The index set of K_p information bits corresponding to the most reliable bit channels and $N_p - K_p$ frozen bits related to the least reliable bit channels are denoted by \mathbb{A} and \mathbb{A}^c , respectively.

Polar codes achieve the channel capacity asymptotically in code length when decoded using the successive cancellation (SC) decoding algorithm, which sequentially estimates the bits \hat{u}_i based on the decoder's input vector \mathbf{y} and the previously estimated bits, specifically, \hat{u}_0 to \hat{u}_{i-1} . The error correction performance of SC decoding when applied to codes with practical length can be improved by using a successive cancellation list (SCL) decoder. The SCL algorithm estimates a bit considering both of its possible values 0 and 1. At every estimation, the number of codeword candidates doubles. In order to limit the increase in the complexity, only a set of L_p codeword candidates is kept in memory at any given time. Therefore, after every estimation, half of the candidates are discarded according to the metric associated with them. [13] shows that concatenating a polar code with a cyclic redundancy code of length ℓ_{crc} can significantly improve the performance of the SCL decoder.

A primary concern regarding polar codes is the high decoding latency of the SC decoding algorithms. In particular, the sequential nature of the SC and SCL decoders significantly reduces the achievable throughput. To avoid poor latency, researchers have recently proposed fast SC decoding algorithms in which many of the computations are parallelized [2–6]. These algorithms are mainly based on identifying certain types of special nodes in the binary tree of the polar codes whose decoding can be performed in parallel rather than following the sequential SC schedule. Moreover, these algorithms reduce or even eliminate the complexity of many check-node operations.

In order to construct a polar code, it is necessary to identify the index set \mathbb{A} based on the reliability order of the polarized bit channels. The reliability order for SC decoders at a given signal-to-noise ratio (SNR) can be found via various methods such as Monte Carlo simulation, Gaussian approximation, and density evolution. In this paper, we use the discretized density evolution (DDE) method in addition

to the Monte Carlo approach in order to find the reliability order of polar codes. It is worth mentioning that we use fast SCL (FSCL) decoding for polar codes in all the simulations discussed in this paper.

2.3 Coded Modulation

Coded modulation (CM) refers to a system where the generated codewords at the transmitter are combined with a higher order modulator with more than one bit per symbol to increase the spectral efficiency (SE). Widely used examples of CM systems are MLC [16, 17]. Let $(c_0, c_1, \dots, c_{m-1})$ be the binary representation of symbols in a 2^m constellation. In traditional MLC, for each bit-level c_i , there is a corresponding equivalent binary bit channel that is protected by an individual binary component code \mathcal{C}_i with rate R_i . Decoding the component codes in MLC can be performed based on multi-stage decoding (MSD), where the component code \mathcal{C}_i is decoded given the results of previously decoded component codes (bit levels). In BICM, unlike MLC, all the bit levels are protected with a single component code. The coded bits are interleaved and grouped into blocks of m bits $(c_0, c_1, \dots, c_{m-1})$ and the corresponding constellation symbol represented by $(c_0, c_1, \dots, c_{m-1})$ is transmitted through the channel.

2.4 PCS

Distributing inputs in an additive white Gaussian noise (AWGN) channel in a uniform manner is less power-efficient than distributing them in a Gaussian manner. CM with PCS uses a non-uniform distribution on transmitted symbols to approximate the Gaussian distribution and overcome the shaping gap. Essentially, the long-standing problem of combining both shaping and coding to obtain capacity-achieving performance was resolved in [14], where a new architecture known as probabilistic amplitude shaping (PAS) was proposed. In PAS architecture, a fraction of the input bits go through a distribution matcher (DM) block, and the rest are directly passed to the output with uniform distribution. Performing systematic coding after the shaping block would prevent the shaped bits from becoming distorted by the FEC block. Therefore, at the symbol mapper, the output bits from the DM can be directly mapped to amplitude bit locations; the unshaped bits and the redundant FEC parity bits are mapped to sign bit locations. The PAS architecture provides remarkable flexibility in terms

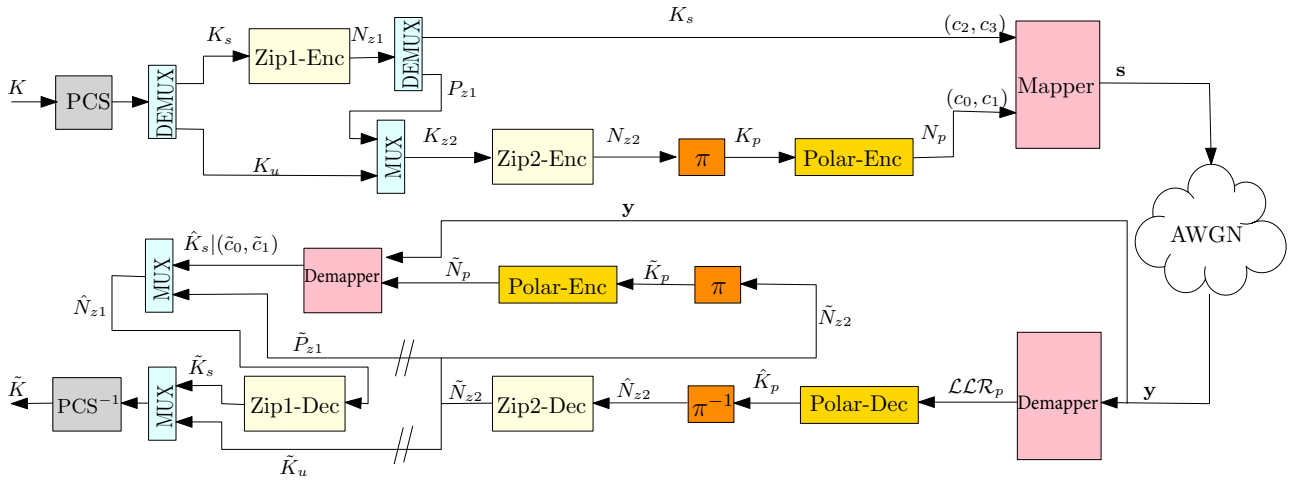


Figure 1 Proposed MLC structure based on concatenated polar-zipper codes

of SE, especially when a fixed FEC scheme is intended to perform in various modes. In our simulation results, we apply PCS based on the PAS architecture.

3 Proposed CM Schemes with Concatenated Polar-Zipper Codes

3.1 MLC

Our proposed MLC scheme is shown in Figure 1. In our MLC scheme, we use a concatenated polar-zipper code to protect the sign bits and use only a hard-decision zipper code to protect the amplitude bits. A set partitioning mapping consistent with the proposed MLC structure is used. The output bit stream of the PCS block consists of shaped bits (K_s) and unshaped bits (K_u). The shaped bits are encoded with zipper encoder 1. Then, the parity bits P_{z1} of zipper code 1 plus the unshaped bits are encoded with zipper encoder 2, while the shaped bits K_s directly go to the symbol mapper. The output bits (N_{z2}) of zipper code 2 are interleaved and passed to a polar encoder. In the mapper, polar codewords (N_p) and shaped bits are mapped to the sign and amplitude bit locations of the constellation symbols, respectively.

At the receiver, MSD is used to decode the noisy codewords. In the first stage, channel log likelihood ratios (LLRs) denoted by \mathcal{LLR}_p corresponding to the sign bits are calculated and passed to the polar decoder. The decoded messages \hat{K}_p of the polar code are de-interleaved and decoded by the corresponding zipper decoder. The error-free zipper codewords \hat{N}_{z2} are used to regenerate error-free

polar codewords \hat{N}_p . Using the channel output y and the knowledge of the estimated sign bits \hat{N}_p , the shaped bits \hat{K}_s are estimated and, after multiplexing with the error-free parity bits \hat{P}_{z1} , are fed to the zipper decoder corresponding to zipper code 1. Finally, the error-free shaped bits \hat{K}_s and the error-free unshaped bits \hat{K}_u are combined through a multiplexer and passed to the de-shaping block PCS^{-1} .

A key feature of our proposed MLC scheme is that, instead of directly using the polar decoder's output bits, the output of hard-decision FEC 2 (which is almost free of error) is used within MSD steps (by passing through the polar encoder) to obtain the LLRs corresponding to the amplitude bits. As a result, there is no need for another interleaver or de-interleaver after the zipper 1 encoder or before the zipper 1 decoder.

The design of the polar code in our MLC scheme is performed using DDE. The target is to achieve the input BER threshold of the zipper code at the lowest possible SNR.

3.2 BICM

Our proposed BICM scheme is shown in Figure 2a. The general structure of this scheme is similar to conventional BICM systems. In our BICM scheme, the output of the shaping block is used to generate the coded bits corresponding to the concatenated polar-zipper codes. The resulting bits are mapped to the constellation symbols and transmitted through channels. Binary representation of symbols follows the Gray labeling criteria. At the receiver, the LLRs corresponding to each bit level are calculated and fed into the polar-zipper decoder module. Finally, the PCS^{-1} block is used to recover the uniformly distributed bits.

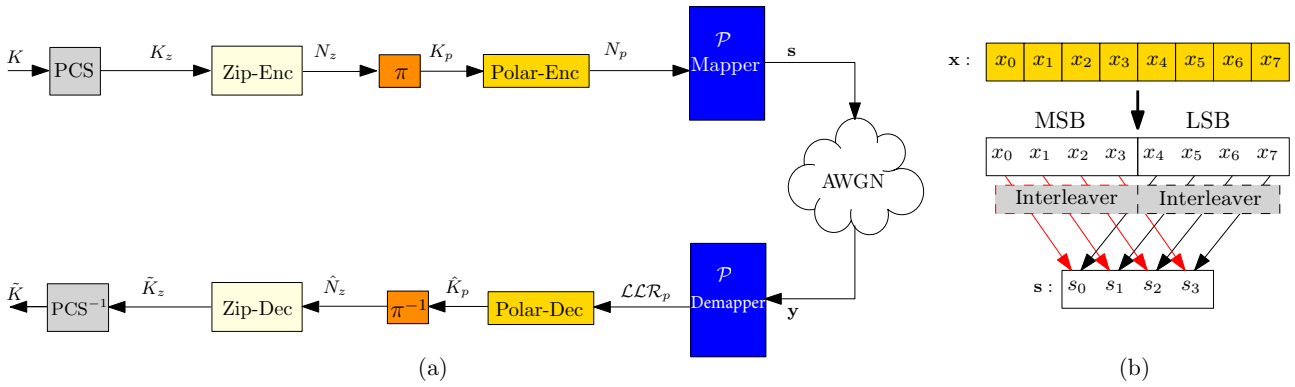


Figure 2 (a) Proposed BICM structure based on concatenated polar-zipper codes; (b) Example of compound polar mapping with a random interleaver over 4PAM. In our proposed scheme, the random interleaver is removed and the polar code is designed without the independence assumption.

One of the critical components in our BICM scheme is the mapping between the polar codeword bits and the bit levels of the 2^m constellation. [18] proposes a compound mapping for polar codes. Figure 2b demonstrates an example of this proposal for a polar code of length 8 over 4PAM. As shown, the first and second halves of the polar codewords are mapped to the most significant bit (MSB) and least significant bit (LSB) positions of different symbols, respectively. In addition, a random interleaver is used to break the dependency of the LLRs going through the check nodes (CNs) and variable nodes (VNs) at the first level of the polar binary tree during decoding.

In our BICM scheme, the random interleaver does not exist, and the MSBs and LSBs are directly mapped to the first and second halves of the polar codewords, respectively. This introduces a correlation among the LLRs going through the same CNs and VNs at the first level of the polar binary tree. We utilize the aforementioned correlation among LLRs to design a polar code that is matched with the proposed mapping. We have observed that the polar code designed based on the proposed correlated mapping outperforms the one designed according to the method of [18]. This is why removing the random interleaver is of practical interest for reducing complexity and latency.

The design of the polar code in our BICM scheme is performed using the Monte Carlo-based approach.

4 Simulation Results

In this section, we investigate the performance of our proposed CM schemes designed based on concatenated polar-zipper codes. We consider transmission through AWGN channels over 16QAM. The simulation parameters and the specifications of polar-zipper codes are summarized

Table 1 Summary of the simulation parameters for 16QAM

Throughput	800G
Baud rate	195
SE	2.23

Table 2 Summary of polar and zipper code specifications in different CM schemes

Scheme	Zipper 1	Zipper 2	Polar	π
	OH% l_{max}	OH% l_{max}	OH% (N_p, K_p, L_p)	Required Memory
MLC	1.49% 4	1.79% 5	61% (8192,5100,8)	1.4M bits
BICM	1.50% 5	N/A N/A	23% (8192,6656,8)	2.0M bits

in Table 1 and Table 2, respectively. We consider the scenario where the throughput, baud rate, and SE are 800G, 195, and 2.23, respectively.

For the proposed MLC scheme in Figure 1, the OH of zipper 1, zipper 2, and polar codes are set to 1.49%, 1.50%, and 61%, respectively. Also, the maximum number of window decoding iterations for zipper 1 and zipper 2 are 4 and 5, respectively, and the list size of the FSCL polar decoder is 8. Regarding the BICM scheme in Figure 2, the OH of zipper and polar codes are 1.50% and 23%, respectively. After fixing the FEC rates, the PCS rate is adjusted such that the desired SE is achieved. While both schemes are generally close in performance, the MLC scheme slightly outperforms the BICM one.

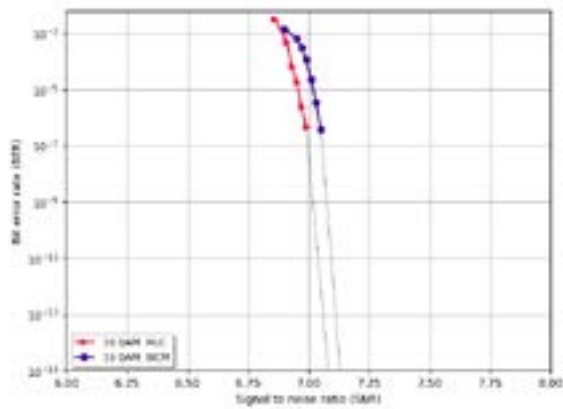


Figure 3 Simulation results of the proposed MLC and BICM schemes based on concatenated polar-zipper codes

5 Conclusion

In this paper, we have proposed two CM schemes, namely, MLC and BICM. For these schemes, we have designed concatenated FEC structures in which the polar and zipper codes are considered as the inner and outer components. In our MLC scheme, the amplitude bits are protected by only a hard-decision zipper code, and only a fraction of bits are protected by a soft polar code. As such, this scheme is of practical interest for low-power applications. For our BICM scheme, we have proposed a bit-to-symbol mapping whose key feature is that no random interleaver is needed. As such, this mapping reduces complexity and latency, while also potentially improving the BER performance.

References

- [1] E. Arikan, "Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels," *IEEE Trans. Inf. Theory*, vol. 55, no. 7, pp. 3051–3073, 2009.
- [2] G. Sarkis, P. Giard, A. Vardy, C. Thibault, and W. J. Gross, "Fast polar decoders: Algorithm and implementation," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 5, pp. 946–957, May 2014.
- [3] G. Sarkis, P. Giard, A. Vardy, C. Thibault, and W. J. Gross, "Fast list decoders for polar codes," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 2, pp. 318–328, Feb. 2016.
- [4] M. Hanif and M. Ardakani, "Fast successive-cancellation decoding of polar codes: Identification and decoding of new nodes," *IEEE Commun. Lett.*, vol. 21, no. 11, pp. 2360–2363, Nov. 2017.
- [5] S. A. Hashemi, C. Condo, and W. J. Gross, "Fast and flexible successive-cancellation list decoders for polar codes," *IEEE Trans. Signal Process.*, vol. 65, no. 21, pp. 5756–5769, 1 Nov. 2017.
- [6] H. Zheng *et al.*, "Threshold-based fast successive-cancellation decoding of polar codes," *IEEE Trans. Commun.*, vol. 69, no. 6, pp. 3541–3555, June 2021.
- [7] A. Süral, E. G. Sezer, Y. Ertuğrul, O. Arikan, and E. Arikan, "Terabits-per-second throughput for polar codes," in *IEEE 30th Inter. Symp. on Personal, Indoor and Mobile Radio Commun.*, 2019, pp. 1–7.
- [8] T. Mehmood, M. P. Yankov, S. Iqbal, and S. Forchhammer, "Flexible multilevel coding With concatenated polar-staircase codes for M-QAM," *IEEE Trans. Commun.*, vol. 69, no. 2, pp. 728–739, Feb. 2021.
- [9] M. Barakatain and F. R. Kschischang, "Low-complexity concatenated LDPC-staircase codes," *J. Lightw. Technol.*, vol. 36, no. 12, pp. 2443–2449, Jun. 15, 2018.

- [10] A. Y. Sukmadji, U. Martínez-Peñas, and F. R. Kschischang, "Zipper codes: Spatially-coupled product-like codes with iterative algebraic decoding," in Proc. Can. Workshop Inf. Theory, Jun. 2019, pp. 1–6.
- [11] B. P. Smith, A. Farhood, A. Hunt, F. R. Kschischang, and J. Lodge, "Staircase codes: FEC for 100 Gb/s OTN," *J. Lightw. Technol.*, vol. 30, no. 1, pp. 110–117, Jan. 2012.
- [12] B. Karimi, M. Barakatain, Y. Hashemi, A. Y. Sukmadji, and C. Pan, "Low latency low overhead zipper codes," patent under review, 2021.
- [13] I. Tal and A. Vardy, "List decoding of polar codes," *IEEE Trans. Inf. Theory*, vol. 61, no. 5, pp. 2213–2226, May 2015.
- [14] G. Bocherer, F. Steiner, and P. Schulte, "Bandwidth efficient and rate-matched low-density parity-check coded modulation," *IEEE Trans. Commun.*, vol. 63, no. 12, pp. 4651–4665, Dec. 2015.
- [15] H. Imai and S. Hiraoka, "A new multilevel coding method using error-correcting codes," *IEEE Trans. Inf. Theory*, vol. IT-23, no. 3, pp. 371–377, May 1977.
- [16] G. Caire, G. Taricco, and E. Biglieri, "Bit-interleaved coded modulation," *IEEE Trans. Inf. Theory*, vol. 44, no. 3, pp. 927–946, May 1998.
- [17] E. Zehavi, "8-PSK trellis codes for a Rayleigh channel," *IEEE Trans. Commun.*, vol. 40, no. 5, pp. 873–884, May 1992.
- [18] H. Mahdavi, M. El-Khamy, J. Lee, and I. Kang, "Polar coding for bit-interleaved coded modulation," *IEEE Trans. Veh. Technol.*, vol. 65, no. 5, pp. 3115–3127, May 2016.



Machine Learning for Quality of Transmission Estimation in Optical Networks

Yvan Pointurier
Paris Research Center

Abstract

Many functions in optical networks rely on the quality of transmission (QoT) of the transported signals being accurately estimated in order to guarantee error-free data transfer. Although many physical models already exist, they typically suffer from the trade-off between speed and accuracy. Additionally, some physical effects are intrinsically difficult to model, and their inputs are not always known accurately due to inaccurate or non-existent monitoring. This is why machine learning (ML) techniques have attracted much attention in the past few years to estimate QoT. We discuss the sources of inaccuracy in QoT estimation, outline a taxonomy for ML-aided QoT estimation, and review and compare recently published ML-aided QoT articles.

Keywords

optical network, machine learning

1 Introduction

Machine learning (ML) has regained popularity in the past few years as a tool capable of solving typically highly nonlinear classification or regression problems for which there is no known analytic solution or for when known solving methods — analytic or heuristic — fail to give an answer in a reasonable time. In the optical communication and networking community, ML has already been extensively leveraged, including:

- 9 surveys within 4 years [1–9]
- 11 workshops or dedicated sessions: at NIST [10, 11] in 2019, OFC conference in 2020 [12], APC conference in 2020 [13], OFC conference in 2021 [14–16], and ECOC conference in 2021 [17–20]
- 3 tutorials at the OFC conference in 2020 [21, 22] and ECOC conference in 2021 [23]

This paper — adapted from [1], which is part of a special issue of the *IEEE/OSA Journal of Optical Communication and Networking* [24] — surveys the body of literature for ML-QoT.

Network designers have long been interested in accurate, fast QoT estimation for services to be established in a future or existing network. Accuracy is important, because modeling errors translate into design margins [25, 26], which in turn translate into wasted capacity [27] or unwanted regeneration [28, 29]. In the case of network planning (either for a new network or for a new service in an existing network), a timescale of seconds for computations is appropriate. In the case of online provisioning, however, computations must be much faster, for instance, well below 1 second per service. One example of online provisioning is restoration.

For network designers, the main QoT metric of interest is service or light path bit error rate (BER), whose value determines whether a service is acceptable and "error free" (whether the BER is below a predefined threshold; equivalently, whether the SNR or Q factor is above a predefined threshold).

This paper is organized as follows: In Section 2, we discuss the two key sources of inaccuracy for QoT estimation, namely, the (physical) model itself and its inputs. In Section 3, we review the taxonomy used in this paper. Our key contribution is Section 4, where we summarize and compare all ML-based QoT estimation papers according to the taxonomy introduced in Section 3.

2 Sources of Inaccuracy in QoT Estimation

Before we discuss why QoT estimation generally remains inaccurate, we need to delve into the main sources of impairments that are accounted for in a QoT metric such as the signal-to-noise ratio (SNR).

SNR can be split into several terms, corresponding to different noise sources. In general, we can write $SNR^{-1} = \sum_k SNR_k^{-1}$, where SNR is the SNR of a service and SNR_k is the contribution of the physical effect indexed by k , for instance:

- linear amplified spontaneous emission (ASE) noise injected by optical amplifiers (in this case, the contribution is called optical SNR, or OSNR);
- nonlinear noise caused by the fiber Kerr effect;
- filtering penalties due to filters being narrower than the signal's bandwidth (and, conversely, node crosstalk, originating from imperfectly filtered signals from adjacent or other ports);
- transponder back-to-back penalty; and
- polarization dependent losses (PDL).

The first three of these effects have been modeled with ML and are discussed further in this paper. Transponder back-to-back penalty is difficult to estimate through ML because it depends on each pair of transceivers and it is usually measured. PDL depends on sources that are external to and independent from the telecommunication infrastructure. These sources are inherently impossible to monitor, and therefore cannot be used as features in an ML framework. Consequently, PDL is difficult to model through ML, and static (e.g., worst-case) margins are often used.

2.1 Trade-off Between Accuracy and Speed

Many physical, non-ML-based models for QoT estimation exist. While there are accurate models for each effect, the speed bottleneck is the modeling of nonlinear effects, for which methods range from the Split Step Fourier (SSF) method to analytic models. The SSF method is highly accurate and versatile, as it can address complex scenarios including the mix of non-coherent and coherent signals in networks with dispersion management. However, SSF is very slow — minutes or more per service.

The (coherent) Gaussian Noise model [30] is much faster but accurate only within its application domain, which does not include, for instance, lines with dispersion management. An approximation of the coherent Gaussian Noise model is the incoherent Gaussian Noise model [30], which is faster but less accurate than the coherent version. Other variations also exist [31, 32], and experimental validations of the Gaussian Noise model can be found in [33–35].

QoT estimation through physical modeling is still a very active field. Its ongoing goal is to capture more effects applicable to more diverse scenarios, such as modeling stimulated Raman scattering, which can no longer be neglected in wideband C or C+L systems [36].

2.2 Inaccurate Inputs

An extremely accurate QoT model (based on physics or ML) is practically useless if its inputs or parameters are not known accurately. The concept of "garbage in, garbage out" holds true in this case — how can we model QoT when the type of fiber may not even be known? [37] It holds to reason, then, that the inputs themselves are a source of inaccuracy in QoT estimation. As mentioned earlier, SNR can be split into several terms (OSNR, nonlinear penalty, etc.), whose inputs may be subject to uncertainty. In Figure 1, we review sources of input or parameter uncertainty with their respective impact on QoT estimation.

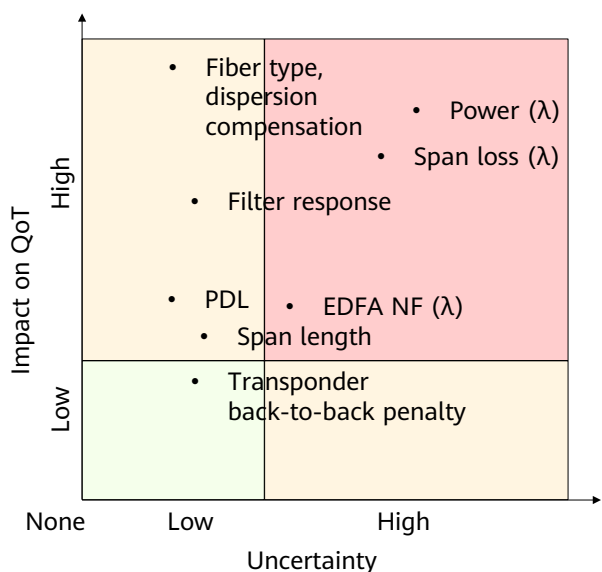


Figure 1 Inaccuracy of typical QoT inputs

Linear noise is typically the main impairment in optical networks. It is driven by per-channel power at the input

of each span, which is not monitored due to a lack of optical spectrum analyzers in amplifiers deployed on current networks. Input powers are in turn tightly related to per-channel EDFA gain, which is difficult to model and is not monitored. We will discuss this later in Section 4.1.

Span losses include both the attenuation of the fiber itself and lumped losses such as splices (which are more numerous in older fibers) or connector losses. The uncertainty involved in span losses results in improper amplifier settings and increased linear noise. Although span losses can be measured in an in-service network, they are difficult to predict during network design — they can also change with time. Wavelength dependence of the span losses is incurred due to stimulated Raman scattering.

Noise figures are usually known, albeit with limited accuracy or without any gain or wavelength dependence. Note that per-channel power also affects nonlinear effect modeling.

The type of fiber being used and whether dispersion compensation/management is employed strongly affect both nonlinear and linear impairments, and are usually (but not always) well known. Span losses (through the span input power) and fiber length both affect nonlinear effect modeling.

Filter responses within wavelength selective switches (WSS) and transponder back-to-back penalties can be discerned from calibration data sheets, but they may vary slightly from component to component and due to aging. Filter degradation has a greater impact on services whose baudrate is close to the filter width. Ultimately, however, the typical PDL of a component is usually provided in data sheets.

2.3 Discussion

Figure 1 is as generic as possible, and strongly depends on the exact network/scenario under consideration: information about the infrastructure (e.g., fiber type) may be lost over time, mistyped in the network management tool, or unavailable by design in the case of disaggregation/alien wavelengths. Similarly, calibration data for filters or transponders may or may not be available, depending on factors such as when the network was built, the equipment vendor, and the operator.

Overall, the three bottlenecks for QoT estimation with traditional (physical) models are currently:

- lack of accurate physical models for some of the components (amplifiers in particular);

- speed of the models for some of the impairments; and
- dependence on uncertain parameters or inputs.

ML can mitigate each of these bottlenecks. Components that are difficult to model can be learned, as too can physical effects that are complex/time-consuming to model. Indeed, many ML techniques have low computational complexity when performing the estimation. This is because the computational complexity is moved to the training phase, which is usually performed offline. For uncertain inputs that cannot be directly monitored, ML can be used to infer them.

Furthermore, ML can also be applied to predict trends of physical phenomena [38]. Such trends can be leveraged to predict the evolution of the parameters or inputs used in traditional or ML-based models, thereby enhancing the models' output accuracy. However, when parameter or input variations are caused by external sources independent from the physics of the considered system, as is the case with PDL, ML will likely be less useful.

3 Taxonomy

This section introduces key concepts used to classify and assess the methods proposed in related literature.

3.1 QoT Models: Classification and Regression

ML models for QoT can be split into two categories: classification and regression. Classification-based formulation typically solves the following problem: will a service to be established (or an existing one if the establishment of a new service is expected to impact other services in a low-margin scenario) have an acceptable QoT (BER below FEC limit)? In this case, QoT is modeled as a binary classifier. Classifiers are sufficient when a network designer needs to know only whether it is possible to establish a service. More often, however, the network designer wants to know the exact value of the QoT. This is needed so that the designer can, for instance, assess the service's robustness to unplanned impairments (how far is the service from the threshold?), or to determine a capacity upgrade of the service — for example, move it from 100 Gb/s PDM-QPSK (quadrature phase-shift keying) to 200 Gb/s PDM-16QAM (quadrature amplitude modulation). In this case, regression techniques are used. Classifiers with more than two classes have been proposed, but they are

uncommon. Each class in such classifiers corresponds to a QoT range, which can be used to coarsely (depending on the number/granularity of the classes) assess the robustness or upgradability of a service.

3.2 Use Cases: Components, Transmission Lines, and Networks

To tackle the problem of estimating QoT for services network-wide, a divide-and-conquer approach can be used. In such an approach, the network can be split into transmission lines (the set of elements between two nodes or ROADMs), and each transmission line can be modeled using ML techniques. Furthermore, transmission lines can be divided into a series of devices (e.g., amplifiers, fibers, WSSs/filters), and each device can also be modeled using ML techniques. Nevertheless, assuming that models even exist for each device, there is no guarantee that their cascade will be accurately modeled. Indeed, small inaccuracies in the behavior of a component accumulate, leading to potentially large end-to-end inaccuracies. This is why network-wide or line-level methods are investigated in addition to the modeling of individual devices. Here we adopt the following criterion to distinguish between line and network studies: Whenever the proposed method uses network-wide data (i.e., the data used to train the model comes from anywhere in the network and not only from the line crossed by a considered service), we classify the paper in the "network" category. We also apply this category to cases where individual lines are considered and trained on synthetic data due to practical reasons, because a realistic scenario would require using training data from already established light paths to model the line. If each transmission line is considered independently one from another with no input coming from any other line, we classify the paper in the "line" category.

3.3 Data Source: Greenfield vs. Brownfield

ML intrinsically relies on a training step, where data is gathered to train or calibrate a model. For instance, data can be:

- generated using a modeling tool known to be accurate but having some other drawback such as speed (e.g., the SSF method or the Gaussian Noise model [30]);
- collected in a lab environment on equipment prior to deployment or on equivalent equipment; or

- collected directly in the field before/during commissioning, or even during network operation (online).

Data collection may be difficult, time-consuming, or costly, and reducing the amount of training is generally desirable. In particular, when a model for a system (device, line, network) has been trained for a specific set of parameters (e.g., a modulation format), adjusting the model for another similar parameter value (e.g., another modulation format) should be possible without having to fully re-train the model. Some of the papers described here leverage transfer learning, which aims to minimize the amount of new data needed to re-train an ML model with no loss of accuracy when changing the value of a key parameter in the considered system.

Because ML relies on a training step, which in turn relies on data collection, most of the ML techniques we describe here apply to in-service "brownfield" line/networks, allowing data to be collected directly from the field to train models. The typical scenario is capacity upgrade, where an operator needs to verify whether the QoT of a new to-be-established service is acceptable, or where the operator wants to determine the maximum capacity that the service can carry given its estimated QoT. Additionally, the operator may want to verify whether the introduction or restoration of a new service has any negative impact on existing services in case the network was designed with low margin [26]. Relying on data from existing services is not the only option; operators can actively probe the network to generate more data ("active learning"). However, some of the presented techniques also apply to lines/networks yet to be physically deployed, where field data is not available — this is the "greenfield" scenario. Estimating the QoT of all services to be established immediately after deployment of a line/network is a typical scenario. An intermediate scenario is when models are trained using a limited amount of data collected immediately after deployment of a line/network for calibration. In this scenario, networks cannot be designed using the model because training data is not yet available. However resource (power, capacity, routing, spectrum, etc.) allocation can still be computed before the network goes live. We consider this to also be a greenfield scenario.

3.4 Learning Type: Data Scientist, Physicist, and Hybrid

In this paper, we introduce a novel type of classification for ML-based algorithms applied to QoT estimation. Consider a physical model f (e.g., the computation of OSNR after a

fiber span and an amplifier) parameterized by θ (e.g., the amplifier's noise figure and gain or the fiber's linear loss parameter, known only from data sheets) to estimate y (e.g., the previously mentioned OSNR) based on inputs or features x (e.g., the launch power of a channel in the fiber): $y=f_{\theta}(x)$.

The traditional approach, which we call "data scientist" (Figure 2a), models a system as a black box $f_{\theta}=f_{\theta}^d$ that has no a priori knowledge of the underlying physics of the system. This does not mean that a data scientist handling the QoT modeling problem should not understand the physics of the system to model. For example, physical skills/insight will play a role when selecting, formatting, or normalizing the features.

The black box (e.g., a neural network) is parameterized (e.g., θ implements the neural weights) and the training step consists of setting θ such that the resulting data model closely approximates the underlying physical model. Most papers we summarize here follow this approach.

In the second approach, which we call "physicist" (Figure 2b), it is assumed that a physical model $f_{\theta}=f_{\theta}^p$ already exists for the model of interest, but that the parameters θ (e.g., the fiber length or linear attenuation) of the system are not known accurately. This approach assumes that f_{θ}^p is correct but that QoT estimation inaccuracies stem from θ . The goal of the training step is then to refine or calibrate θ in order to increase the accuracy of f_{θ}^p when presented with a new input.

In the third approach, which we call "hybrid" (Figure 2c), the system of interest is again modeled as a black box $f_{\theta}=f_{\theta}^d$, as in the data scientist approach (e.g., f_{θ}^d is a neural network), and the traditional inputs are supplemented with the output of the physical model corresponding to the system f^p (e.g., the analytic OSNR calculation for a fiber span). This biases the data scientist model towards the behavior of the physical model, while physical modeling deficiencies are addressed through training on the real system.

4 ML-QoT Estimation

The key features according to the taxonomy presented in Section 3 and the outcomes of the papers surveyed here are summarized in Tables 1–4. Each table corresponds to one of the subsections in this section (namely, amplifiers modeling, line modeling with regression, network modeling with classification, and network modeling with regression). Following the discussion in Section 3.3, we left the "Data Source" column empty for the transfer learning-related

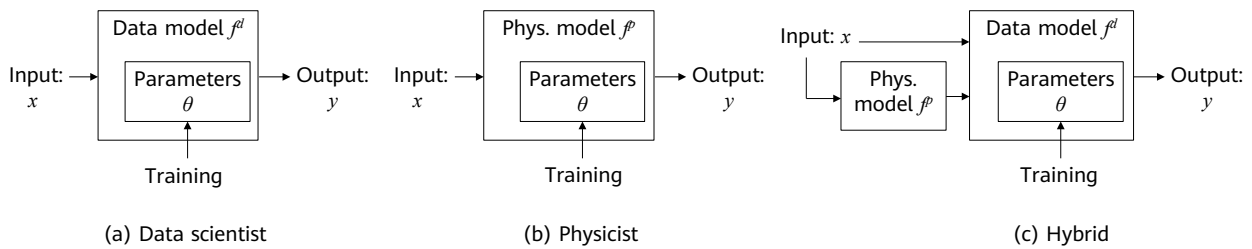


Figure 2 Machine learning approaches for QoT estimation

techniques. In the tables, we indicate whether results were obtained experimentally ("Exp.") or through simulation ("Sim.") in the "Type of Study" column. We also list the features needed to estimate the QoT metric in the "Inputs" column. All algorithms are supervised and rely on training data, which can be obtained through monitoring on a live network, emulation/calibration (e.g., from lab experiments), or simulation (e.g., using a physical model).

4.1 Amplifiers Modeling

The first set of papers (Table 1) aim to model the behavior of amplifiers: a single EDFA ([41–44]), Raman amplification with an EDFA [45], and a single Raman amplifier (along with a 100 km fiber span) over the C+L bands [46].

Modeling of amplifiers is important, because most systems are OSNR-limited; specifically, the main impairment is the ASE noise originating from amplification. The amount (power variance) of noise generated by an amplifier is deceptively simple and given by $P_{ASE} = hfB(GF-1)$, where h is Planck's constant, f is the central frequency of the signal, B is the bandwidth in which the noise is considered, G is the amplifier gain, and F is its noise figure. In reality, however, EDFA gains are wavelength dependent and the shape of the gain nonlinearly depends on the input power spectrum of the EDFA. Specifically, it depends on which input channels are lit, and on the power of each of the input channels. This is caused by effects such as spectral hole burning, which is difficult to model and affects amplifier gain at the short wavelengths in the C band [47]. EDFAs are therefore good candidates for ML modeling, as too are Raman amplifiers that are modeled using a set of ordinary differential equations [48].

The papers mentioned earlier in this section essentially follow the same concept: given the input power spectrum of an amplifier, estimate the gain per channel or equivalently the output power spectrum using a trained neural network (for instance, trained in the lab) prior to deployment. With

the exception of [44], which uses the "hybrid" approach, all amplifier modeling papers use the "data scientist" approach. In [44], a neural network is fed with the outputs of an EDFA analytic model in addition to the standard modeling inputs. Although this technique ultimately does not improve modeling accuracy, it improves the training time and reduces the amount of training.

Each study was done for a single amplifier with a single setting (e.g., the amplifier gain or total input power) except for [45], which additionally outputs the amplifier's settings to achieve a given response. In [49], as related work, the problem of finding an amplifier's settings to achieve a target gain shape is put into the broader context of inverse system design. It uses an ML framework to learn what parameters yield a desired output for a given system, such as an optical amplifier.

Solving the problem for a single setting implies that training may have to be done for each amplifier's settings. The reported results — an error (RMSE or maximum) typically below 0.2 dB — should be considered in light of the estimation framework being applied. In a line with 10 or 20 such amplifiers, errors may accumulate to several dB in optical spectrum and then in SNR.

4.2 Line Modeling: Regression

We now consider the modeling of an individual optical point-to-point line (Table 2) [50–56]. All papers except [50], which leverages Gaussian Process Regression (GPR), rely on neural networks. And except [53], which uses data from the GN model for training (hybrid approach), all papers follow a "data scientist" approach.

In [50, 52], the greenfield scenario is considered: [50] shows that training the GPR model on synthetic data enables accurate BER estimation in deployed systems, while [52] assumes that a line model is trained prior to deployment or at commissioning time based on measurements on the real line. Both [51] (in the very special case of a submarine link

subject to strict power supply constraints) and [52] model only linear effects. This is because, in [51], operation far from the nonlinear regime is assumed, whereas in [52], it is assumed that nonlinear effects can be accurately modeled (e.g., using the Gaussian Noise model) and that most uncertainty on QoT modeling comes from the lack of an accurate model for the amplifiers, as discussed in Section 4.1. In [52], modeling experiments are carried out on a cascade of amplifiers without any fibers, whose losses are emulated with variable optical amplifiers. In all cases, estimation error of a fraction of a dB is reported. Papers [54, 55] further propose transfer learning techniques to adjust a neural network trained on one line to another line with little retraining in order to estimate the OSNR [54] or Q [55].

4.3 Network Modeling: Classification

The following set of papers (Table 3) adopt a classification framework, generally to estimate whether a service has an acceptable QoT [57–62] and, in the case of [63], to coarsely estimate the value of the OSNR (because only ASE noise is modeled) through binning using a four-class classifier. For the sake of completeness, classification is applied to multicast services in [58], generalized to multicore networks in [64], and to multi-slice networks with distributed controllers in [65]. [66] further leverages the work from [39] for improved resource allocation.

Typical classification accuracy is 90% and reaches 99% in several of the papers. A problem with classification is the lack of training data for the "unacceptable BER" class. This is because operators typically do not try to establish services they anticipate will fail, or at least target a very small blocking rate. This problem biases the classifier and is usually solved by generating data (services) with unacceptable BER by simulation. Such data generation is not practical in live networks, where forcing the collection of samples in that class could instead be done through probing, including establishing services with poor BER as in [39, 59]. In the case of probing, sending new signals in a functioning network must not disturb network operations; the cost of probing equipment should also be considered.

In [62], the amount of training data for the binary classifier proposed in [39] is reduced through transfer learning when the classifier is applied to a new line or network. This approach obtains similar accuracy on the target without requiring full retraining of the ML model. In [67], the transfer learning technique is compared with the

probing-based active learning technique already mentioned earlier [59] in terms of samples to collect. Compared with transfer learning, active learning requires fewer samples, but those that it does require are targeted samples that may be more difficult or costly to collect. All papers rely exclusively on simulations, based on synthetic data sets generated through one of the QoT analytic models introduced in Section 2.1, such as the GN model.

It should be noted that only a remarkably small quantity of features are required to achieve high classification accuracy. In particular, none of the features needed to model nonlinear effects is present in some of the papers. For example, per-channel or at least total input span power is not present. Neither is the length of each span. Because nonlinear effects occur at the beginning of the spans, if all spans are sufficiently long, then nonlinear effects are similar for all spans and the exact lengths are not needed. However, a high diversity in span lengths would yield a high diversity in nonlinear impairments. Additionally, span lengths drive amplification noise. The lack of such features suggests that OSNR-limited scenarios are considered, or that even higher classification accuracy could be achieved with additional features. This discussion can be generalized to the regression papers, not all of which use the features that are needed to physically model nonlinear effects.

4.4 Network Modeling: Regression

In the last set of papers (Table 4; see also Table 3 for [61]) [29, 61, 68–84], network-wide information is leveraged to train and estimate the QoT of a service in the brownfield/upgrade scenario with one exception: [78], which also applies to greenfield. Those techniques are highly diverse: 9 papers use a data scientist approach, 7 a physicist approach, and 4 a hybrid approach, using diverse techniques such as neural networks, optimization through filtering, gradient descent, least squares, or SVM regression. 11 papers report simulations and 7 report experiments.

The physicist approach is used in [29, 69, 71, 76, 77, 83, 84] to learn the parameters of a physical model or to refine them in case those inputs are not accurately known. In particular, [77] focuses on learning EDFA gain ripple penalties from end-to-end measurements (as mentioned in Section 4.1, such a technique can be cast in the more general framework of inverse system design), while [76] additionally learns filtering penalties and [83, 84] connector losses. The other 3 papers [29, 69, 71] deal with end-to-end QoT as a whole.

Note that [71] also uses the data scientist approach. The problem of handling parameter uncertainty is also tackled in [78] using the hybrid approach, but without trying to refine the parameters.

The hybrid approach is used in [72, 73, 78, 81], where artificial neural networks are fed with not only network- and signal-related features as in other works, but also the output of a modeling tool (e.g., the Gaussian Noise model) optionally complemented with monitoring data. Paper [72] additionally considers networks with mixed types of fibers, whereas [78] leverages information gained from a network to apply to another network, without relying on transfer learning.

5 Conclusion

QoT estimation is just one of many optical networking fields in which ML has been applied. In only four years, dozens of articles have been published on the specific topic of ML-QoT estimation.

This field is extremely diverse, as exemplified by the variety of the techniques/ML tools that are used and the targeted application — from individual components modeling to full network modeling, from modeling a single physical effect to modeling end-to-end BER, and from networks still at the design stage (greenfield) to networks already deployed and waiting for an upgrade (brownfield).

ML proves to be useful when traditional modeling is difficult or slow, which is the case for some key optical networking components such as optical amplifiers, as well as full lines (which cascade said components) and consequently whole networks. In addition to replacing physical models through a black box approach, ML can also complement existing physical models through refinement of the model parameters. In terms of accuracy and speed, the results reported in the related literature are very encouraging. Of course, as in any ML-related research, the quality of the underlying data is essential.

In particular, ML relies on training, which is often specific to components or lines. Solutions include training in the lab (at the expense of real field experience and possibly accuracy), online (which limits the use case to "brownfield" or capacity upgrade, but is likely more accurate through reliance on actual, monitored data), or through retraining a model already trained in the lab as with transfer learning techniques.

Researchers have to therefore find the right trade-off between synthetic data generation — where ML may simply end up learning the algorithm used to generate the data — and experimental studies, which are limited in scope due to the difficulty and cost of emulating a wide range of scenarios.

Acknowledgments

Thanks to Gabriel Charlet for the insightful discussions.

References

- [1] Y. Pointurier, "Machine learning techniques for quality of transmission estimation in optical networks," *IEEE/OSA J. Opt. Commun. Netw.* 13, B60–B71 (2021).
- [2] J. Mata, I. de Miguel, R. J. Durán, N. Merayo, S. K. Singh, A. Jukan, and M. Chamania, "Artificial intelligence (AI) methods in optical networks: A comprehensive survey," *Elsevier Opt. Switch. Netw.* 28, 43–57 (2018).
- [3] D. Rafique and L. Velasco, "Machine learning for network automation: Overview architecture, and applications," *IEEE/OSA J. Opt. Commun. Netw.* 10, D126–D143 (2018).
- [4] F. Musumeci, C. Rottondi, A. Nag, I. Macaluso, D. Zibar, M. Ruffini, and M. Tornatore, "An overview on application of machine learning techniques in optical networks," *IEEE Commun. Surv. & Tutorials* 21, 1383–1408 (2019).
- [5] M. A. Amirabadi, "A survey on machine learning for optical communication [machine learning view]," [arXiv:1909.05148 \[eess.SP\]](https://arxiv.org/abs/1909.05148) (2019).
- [6] R. Gao, L. Liu, X. Liu, H. Lun, L. Yi, W. Hu, and Q. Zhuge, "An overview of ML-based applications for next generation optical networks," *Sci. China Inf. Sci.* 63 (2020).
- [7] X. Liu, H. Lun, M. Fu, Y. Fan, L. Yi, W. Hu, and Q. Zhuge, "AI-based modeling and monitoring techniques for future intelligent elastic optical networks," *MDPI Appl. Sci.* p. 363 (2020).
- [8] R. M. Morais, "On the suitability, requisites, and challenges of machine learning [Invited]," *J. Opt. Commun. Netw.* 13, A1–A12 (2021).
- [9] J. W. Nevin, S. Nallaperuma, N. A. Shevchenko, X. Li, Md. S. Faruk, and S. J. Savory, "Machine learning for optical fiber communication systems: An introduction and overview," *APL Photonics* 6, 121101 (2021); <https://doi.org/10.1063/5.0070838>
- [10] J. Gordon, A. Battou, M. Majurski, D. Kilper, U. Celine, D. Zibar, M. Tornatore, J. Pedro, J. Simsarian, and J. Westdorp, "Summary: Workshop on machine learning for optical communication systems," <https://doi.org/10.6028/NIST.SP.2100-04>
- [11] J. Gordon, A. Battou, and D. Kilper, "Workshop on machine learning for optical communication systems: A summary," in *Proc. IEEE/OSA Optical Fiber Communication Conference (OFC)*, (San Diego, CA, USA, 2020), p. M3E.3.
- [12] "The role of machine learning for the next-generation of optical communication systems and networks," (2020). *IEEE/OSA Optical Fiber Communication Conference (OFC)*, sessions M3E and M4E.
- [13] "AI-based optics," (2020). *Asia Communications and Photonics Conference (ACP)*, workshop 1.
- [14] "Machine learning in optical networks," *IEEE/OSA Optical Fiber Communication Conference (OFC)*, session Tu1G, 2021.
- [15] "Machine learning for failure management and performance estimation / prediction," *IEEE/OSA Optical Fiber Communication Conference (OFC)*, session Th4J, 2021.
- [16] "Machine learning and data analytics for autonomous network operation," *IEEE/OSA Optical Fiber Communication Conference (OFC)*, session F2G, 2021.
- [17] "How machine learning can revolutionize optical fiber communications?" *European Conference on Optical Communication (ECOC)*, workshop Mo1B, 2021.
- [18] "Machine learning I," *European Conference on Optical Communication (ECOC)*, session Tu1C2, 2021.
- [19] "Machine learning II," *European Conference on Optical Communication (ECOC)*, session Tu2C2, 2021.
- [20] "Machine learning for optical networks," *European Conference on Optical Communication (ECOC)*, session Th2C2, 2021.
- [21] R. Morais, "Machine learning in multi-layer optical networks: Why and how," in *Proc. IEEE/OSA Optical Fiber Communication Conference (OFC)*, (San Diego, CA, USA, 2020), p. M1B.1.
- [22] F. N. Khan, Q. Fan, J. Lu, G. Zhou, C. Lu, and A. Pak Tao Lau, "Applications of machine learning in optical communications and networks," in *Proc. IEEE/OSA Optical Fiber Communication Conference (OFC)*, (San Diego, CA, USA, 2020), p. M1G.5.
- [23] C. Tremblay, "Neural networks for estimating

- lightpath QoT in optical communications," European Conference on Optical Communication (ECOC), tutorial Th1C1.1, 2021.
- [24] Y. Pointurier, J. Pesic, C. Rottondi, and L. Velasco, "Introduction to the JOCN special issue on machine learning applied to QoT estimation in optical networks," *IEEE/OSA J. Opt. Commun. Netw.* 13, MLAQ1–MLAQ2 (2021).
- [25] J.-L. Augé, "Can we use flexible transponders to reduce margins?" in *Proc. IEEE/OSA Optical Fiber Communication Conference (OFC)*, (Anaheim, CA, USA, 2013), p. OTu2A.1.
- [26] Y. Pointurier, "Design of low-margin optical networks," *IEEE/OSA J. Opt. Commun. Netw.* 9, A9–A17 (2017).
- [27] S. J. Savory, R. J. Vincent, and D. J. Ives, "Design considerations for low-margin elastic optical networks in the nonlinear regime," *IEEE/OSA J. Opt. Commun. Netw.* 11, C76–C85 (2019).
- [28] I. Sartzetakis, K. Christodouloupoulos, C. P. Tsekrekos, D. Syvridis, and E. Varvarigos, "Quality of transmission estimation in WDM and elastic optical networks accounting for space–spectrum dependencies," *IEEE/OSA J. Opt. Commun. Netw.* 8, 676–688 (2016).
- [29] M. Bouda, S. Oda, O. Vassilieva, M. Miyabe, S. Yoshida, T. Katagiri, Y. Aoki, T. Hoshida, and T. Ikeuchi, "Accurate prediction of quality of transmission based on a dynamically configurable optical impairment model," *IEEE/OSA J. Opt. Commun. Netw.* 10, A102–A109 (2018).
- [30] P. Poggiolini, G. Bosco, A. Carena, V. Curri, Y. Jiang, and F. Forghieri, "The GN-model of fiber non-linear propagation and its applications," *J. Light. Technol.* 32, 694–721 (2014).
- [31] R. Dar, M. Feder, A. Mecozzi, and M. Shtaif, "Properties of nonlinear noise in long, dispersion-uncompensated fiber links," *Opt. Express* 21, 25685–25699 (2013).
- [32] P. Serena and A. Bononi, "A time-domain extended Gaussian noise model," *J. Light. Technol.* 33, 1459–1472 (2015).
- [33] J. Lu, G. Zhou, Q. Fan, D. Zeng, C. Guo, L. Lu, J. Li, C. Xie, C. Lu, F. N. Khan, and A. Pak Tao Lau, "Performance comparisons between machine learning and analytical models for quality of transmission estimation in wavelength-division-multiplexed systems [Invited]," *IEEE/OSA J. Opt. Commun. Netw.* 12, B35–B44 (2021).
- [34] A. Ferrari, M. Filer, K. Balasubramanian, Y. Yin, E. Le Rouzic, J. Kandrát, G. Grammel, G. Galimberti, and V. Curri, "GNPy: An open source application for physical layer aware open optical networks," *IEEE/OSA J. Opt. Commun. Netw.* 12, C31–C40 (2020).
- [35] A. Ferrari, K. Balasubramanian, M. Filer, J. Kandrát, Y. Yin, G. Grammel, G. Galimberti, E. Le Rouzic, and V. Curri, "Softwarized optical transport QoT in production optical network: A brownfield validation," in *Proc. European Conference on Optical Communication (ECOC)*, (2020).
- [36] D. Semrau, R. I. Killey, and P. Bayvel, "The Gaussian noise model in the presence of inter-channel stimulated Raman scattering," *J. Light. Technol.* 36, 3046–3055 (2018).
- [37] E. Seve, J. Pesic, C. Delezoide, A. Giorgetti, A. Sgambelluri, N. Sambo, S. Bigo, and Y. Pointurier, "Automated fiber type identification in SDN-enabled optical networks," *J. Light. Technol.* 37, 1724–1731 (2019).
- [38] C. Wang, S. Fu, Z. Xiao, M. Tang, and D. Liu, "Long short-term memory neural network (LSTM-NN) enabled accurate optical signal-to-noise ratio (OSNR) monitoring," *J. Light. Technol.* 37, 4140–4146 (2019).
- [39] C. Rottondi, L. Barletta, A. Giusti, and M. Tornatore, "Machine-learning method for quality of transmission prediction of unestablished lightpaths," *IEEE/OSA J. Opt. Commun. Netw.* 10, A286–A297 (2018).
- [40] M. Lonardi, J. Pesic, T. Zami, and N. Rossi, "The perks of using machine learning for QoT estimation with uncertain network parameters," in *Proc. OSA Advanced Photonics Congress*, (2020).
- [41] S. Zhu, C. L. Gutterman, W. Mo, Y. Li, G. Zussman, and D. C. Kilper, "Machine learning based prediction of erbium-doped fiber WDM line amplifier gain spectra," in *Proc. European Conference on Optical Communication (ECOC)*, (Rome, Italy, 2018).
- [42] Y. You, Z. Jiang, and C. Janz, "Machine learning-based EDFA gain model," in *Proc. European Conference on Optical Communication (ECOC)*, (Rome, Italy, 2018).

- [43] M. Freire, S. Mansfeld, D. Amar, F. Gillet, A. Lavignotte, and C. Lepers, "Predicting optical power excursions in erbium doped fiber amplifiers using neural networks," in Proc. Asia Communications Conference (ACP), (Hangzhou, China, 2018).
- [44] J. Yu, S. Zhu, C. Gutterman, G. Zussman, and D. Kilper, "Machine-learning-based EDFA gain estimation [Invited]," *IEEE/OSA J. Opt. Commun. Netw.* 13, B83–B91 (2021).
- [45] M. Ionescu, "Machine learning for ultrawide bandwidth amplifier configuration," in Proc. International Conference on Transparent Optical Networks (ICTON), (Angers, France, 2019).
- [46] A. M. Rosa Brusin, U. C. De Moura, V. Curri, D. Zibar, and A. Carena, "Introducing load aware neural networks for accurate predictions of Raman amplifiers," *J. Light. Technol.* 38, 6481–6491 (2020).
- [47] M. Bolshtyansky, "Spectral hole burning in erbium-doped fiber amplifiers," *J. Light. Technol.* 21, 1032–1038 (2003).
- [48] J. Bromage, "Raman amplification for fiber communications systems," *J. Light. Technol.* 22, 73–93 (2004).
- [49] D. Zibar, A. M. Rosa Brusin, U. C. De Moura, F. Da Ros, V. Curri, and A. Carena, "Inverse system design using machine learning: The Raman amplifier case," *J. Light. Technol.* 38, 736–753 (2020).
- [50] J. Wass, J. Thrane, M. Piels, R. Jones, and D. Zibar, "Gaussian process regression for WDM system performance prediction," in Proc. IEEE/OSA Optical Fiber Communication Conference (OFC), (Los Angeles, CA, USA, 2017), p. Tu3D.7.
- [51] J. Cho, S. Chandrasekhar, E. Sula, S. Olsson, E. Burrows, G. Raybon, R. Ryf, N. Fontaine, J.-C. Antona, S. Grubb, P. Winzer, and A. Chraplyvy, "Supply-power-constrained cable capacity maximization using multilayer neural networks," *J. Light. Technol.* 38, 3652–3662 (2020).
- [52] A. D'Amico, S. Straullu, A. Nespola, I. Khan, E. London, E. Virgillito, S. Piciaccia, A. Tanzi, G. Galimberti, and V. Curri, "Using machine learning in an open optical line system controller," *IEEE/OSA J. Opt. Commun. Netw.* 12, D84–D99 (2020).
- [53] A. D'Amico, S. Straullu, G. Borraccini, E. London, S. Bottacchi, S. Piciaccia, A. Tanzi, A. Nespola, G. Galimberti, S. Swail, and V. Curri, "Enhancing lightpath QoT computation with machine learning in partially disaggregated optical networks," in *IEEE Open Journal of the Communications Society*, vol. 2, pp. 564–574, 2021, doi: 10.1109/OJCOMS.2021.3066913.
- [54] L. Xia, J. Zhang, S. Hu, M. Zhu, Y. Song, and K. Qiu, "Transfer learning assisted deep neural network for OSNR estimation," *Opt. Express* 27, 19398–19406 (2019).
- [55] J. Yu, W. Mo, Y.-K. Huang, E. Ip, and D. C. Kilper, "Model transfer of QoT prediction in optical networks based on artificial neural networks," *IEEE/OSA J. Opt. Commun. Netw.* 11, C48–C57 (2019).
- [56] S. Kamel, H. Hafermann, D. Le Gac, L. Dos Santos, B. Kégl, Y. Frignac, and G. Charlet, "OSNR prediction for optical links via learned noise figures," in Proc. European Conference on Optical Communication (ECOC), (Bordeaux, France, 2021), We4E.3.
- [57] J. Mata, I. de Miguel, R. J. Durán, J. C. Aguado, N. Merayo, L. Ruiz, P. Fernández, R. M. Lorenzo, and E. J. Abril, "A SVM approach for lightpath QoT estimation in optical transport networks," in Proc. IEEE BigData, (Boston, MA, USA, 2017), pp. 4795–4797.
- [58] T. Panayiotou, S. P. Chatzis, and G. Ellinas, "Performance analysis of a data-driven quality-of-transmission decision approach on a dynamic multicast-capable metro optical network," *IEEE/OSA J. Opt. Commun. Netw.* 9, 98–108 (2017).
- [59] D. Azzimonti, C. Rottondi, and M. Tornatore, "Reducing probes for quality of transmission estimation in optical networks with active learning," *IEEE/OSA J. Opt. Commun. Netw.* 12, A38–A48 (2020).
- [60] S. Aladin, A. V. S. Tran, S. Allogba, and C. Tremblay, "Quality of transmission estimation and short-term performance forecast of lightpaths," *J. Light. Technol.* 38, 2807–2814 (2020).
- [61] R. M. Morais and J. Pedro, "Machine learning models for estimating quality of transmission in DWDM networks," *IEEE/OSA J. Opt. Commun. Netw.* 10, D84–D99 (2018).
- [62] C. Rottondi, R. di Marino, M. Nava, A. Giusti, and

- A. Bianco, "On the benefits of domain adaptation techniques for quality of transmission estimation in optical networks," *IEEE/OSA J. Opt. Commun. Netw.* 13, A34–A43 (2021).
- [63] A. Diaz-Montiel, S. Aladin, C. Tremblay, and M. Ruffini, "Active wavelength load as a feature for QoT estimation based on support vector machine," in *Proc. IEEE International Conference on Communications (ICC)*, (San Diego, CA, USA, 2019).
- [64] T. Panayiotou, G. Savva, B. Shariati, I. Tomkos, and G. Ellinas, "Machine learning for QoT estimation of unseen optical network states," in *Proc. IEEE/OSA Optical Fiber Communication Conference (OFC)*, (San Diego, CA, USA, 2019), p. Tu2E.2.
- [65] T. Panayiotou, G. Savva, I. Tomkos, and G. Ellinas, "Decentralizing machine-learning-based QoT estimation for sliceable optical networks," *IEEE/OSA J. Opt. Commun. Netw.* 12, 146–162 (2020).
- [66] M. Salani, C. Rottondi, and M. Tornatore, "Routing and spectrum assignment integrating machine-learning-based QoT estimation in elastic optical networks," in *Proc. IEEE INFOCOM*, (Paris, France, 2019), pp. 1738–1746.
- [67] D. Azzimonti, C. Rottondi, A. Giusti, M. Tornatore, and A. Bianco, "Comparison of domain adaptation and active learning techniques for quality of transmission estimation with small-sized training datasets," *IEEE/OSA J. Opt. Commun. Netw.* 13, A56–A66 (2021).
- [68] P. Samadi, D. Amar, C. Lepers, M. Lourdiane, and K. Bergman, "Quality of transmission prediction with machine learning for dynamic operation of optical WDM networks," in *Proc. European Conference on Optical Communication (ECOC)*, (Gothenburg, Sweden, 2017).
- [69] E. Seve, J. Pesic, C. Delezoide, S. Bigo, and Y. Pointurier, "Learning process for reducing uncertainties on network parameters and design margins," *IEEE/OSA J. Opt. Commun. Netw.* 10, A298–A306 (2018).
- [70] R. Proietti, X. Chen, K. Zhang, G. Liu, M. Shamsabardeh, A. Castro, L. Velasco, Z. Zhu, and S. J. B. Yoo, "Experimental demonstration of machine-learning-aided QoT estimation in multi-domain elastic optical networks with alien wavelengths," *IEEE/OSA J. Opt. Commun. Netw.* 11, A1–A10 (2019).
- [71] I. Sartzetakis, K. Christodoulopoulos, and E. Varvarigos, "Accurate quality of transmission estimation with machine learning," *IEEE/OSA J. Opt. Commun. Netw.* 11, 140–150 (2019).
- [72] Q. Zhuge, X. Zeng, H. Lun, M. Cai, X. Liu, L. Yi, and W. Hu, "Application of machine learning in fiber nonlinearity modeling and monitoring for elastic optical networks," *J. Light. Technol.* 37, 3055–3063 (2019).
- [73] R. Morais, "Fast and high-precision optical performance evaluation for cognitive optical networks," in *Proc. IEEE/OSA Optical Fiber Communication Conference (OFC)*, (San Diego, CA, USA, 2020), p. Th3D.3.
- [74] J. Pesic, M. Lonardi, N. Rossi, T. Zami, E. Seve, and Y. Pointurier, "How uncertainty on the fiber span lengths influences QoT estimation using machine learning in WDM networks," in *Proc. IEEE/OSA Optical Fiber Communication Conference (OFC)*, (San Diego, CA, USA, 2020), p. Th3D.5.
- [75] Z. Gao, S. Yan, J. Zhang, M. Mascarenhas, R. Nejabati, Y. Ji, and D. Simeonidou, "ANN-based multi-channel QoT-prediction over a 563.4-km field-trial testbed," *J. Light. Technol.* 38, 2646–2655 (2019).
- [76] A. Mahajan, K. Christodoulopoulos, R. Martínez, S. Spadaro, and R. Muñoz, "Modeling EDFA gain ripple and filter penalties with machine learning for accurate QoT estimation," *J. Light. Technol.* 38, 2616–2629 (2020).
- [77] E. Seve, J. Pesic, and Y. Pointurier, "Associating machine-learning and analytical models for QoT estimation: Combining the best of both worlds," *IEEE/OSA J. Opt. Commun. Netw.* 13, C21–C30 (2021).
- [78] I. Khan, M. Bilal, M. Siddiqui, M. Khan, A. Ahmad, M. Shahzad, and V. Curri, "QoT estimation for light-path provisioning in un-seen optical networks using machine learning," in *Proc. International Conference on Transparent Optical Networks (ICTON)*, (2020).
- [79] C.-Y. Liu, X. Chen, R. Proietti, and S. J. B. Yoo, "Performance studies of evolutionary transfer learning for end-to-end QoT estimation in multi-domain optical networks [Invited]," *IEEE/OSA J. Opt. Commun. Netw.* 13, B1–B11 (2021).

- [80] J. Pestic, M. Lonardi, E. Seve, N. Rossi, and T. Zami, "Transfer learning from unbiased training data sets for QoT estimation in WDM networks," in Proc. European Conference on Optical Communication (ECOC), (2020).
- [81] I. Khan, M. Bilal, M. U. Masood, A. D'Amico, and V. Curri, "Lightpath QoT computation in optical networks assisted by transfer learning," *IEEE/OSA J. Opt. Commun. Netw.* 13, B72–B82 (2021).
- [82] M. Ibrahimi, H. Abdollahi, C. Rottondi, A. Giusti, A. Ferrari, V. Curri, and M. Tornatore, "Machine learning regression for QoT estimation of unestablished lightpaths," *IEEE/OSA J. Opt. Commun. Netw.* 13, B92–B101 (2021).
- [83] N. Morette, I. Fernandez de Jauregui Ruiz, and Y. Pointurier, "Leveraging ML-based QoT tool parameter feeding for accurate WDM network performance prediction," in Proc. IEEE/OSA Optical Fiber Communication Conference (OFC), (2021), p. Th4J.4.
- [84] N. Morette, I. Fernandez de Jauregui Ruiz, Hartmut Hafermann, and Y. Pointurier, "On the robustness of a ML-based method for QoT tool parameter refinement in partially loaded networks," in Proc. IEEE/OSA Optical Fiber Communication Conference (OFC), (2022), p. M3F.1.

Table 1 Comparison of ML-QoT estimation frameworks for amplifiers

Paper	Use Case	Data Source	Learning Type	Algorithm	Type of Study	Inputs (Features)	Output	Key Results
[42]	Single EDFA	Green, brown	Data scientist	DNN	Exp.	Per-channel input power	Gain	0.18 dB max. gain estimation error
[43]	Single EDFA	Green, brown	Data scientist	ANN	Sim.	Per-channel input status (on/off)	Optical power spectrum	Error < 0.2 dB in 90% of the cases
[44]	Single EDFA	Green, brown	Hybrid	DNN	Exp.	Per-channel input power; per-channel output power from analytic model	Optical power spectrum	RMSE < 0.2 dB; faster convergence with inputs from analytic model
[45]	Single Raman / EDFA link	Green, brown	Data scientist	DNN	Exp.	Amplifiers' pump powers (which adjust gain and tilt)	Optical power spectrum	Typical error in output power < 0.2 dB (max.: 0.3 dB)
[46]	Single Raman link	Green, brown	Data scientist	ANN	Sim.	Per-subband input power, pump powers	Gain and ASE noise prediction	Max. error < 0.5 dB on gain and noise power

Table 2 Comparison of ML-QoT estimation frameworks for lines: regression

Paper	Use Case	Data Source	Learning Type	Algorithm	Type of Study	Inputs (Features)	Output	Key Results
[50]	Line	Green, brown	Data scientist	GPR	Sim. / Exp.	TX power, number of spans, baudrate, interchannel spacing	BER / Q	Q error < 0.3 dB
[51]	Line	Brown	Data scientist	DNN	Exp.	Per-channel received signal and noise	OSNR	0.5 dB max. OSNR error
[52]	Line	Green, brown	Data scientist	ANN	Exp.	End of line spectrum	OSNR	Error on OSNR < 0.2 dB
[53]	Line	Green, brown	Hybrid	ANN	Exp.	End of line spectrum, signal and ASE power from a model	OSNR	Error on OSNR < 0.2 dB
[54]	Line	See Section 3.3	Data scientist	DNN, transfer learning	Exp.	Amplitude histogram of received samples	OSNR	Retraining test size divided by 5, training time divided by 10
[55]	Line	See Section 3.3	Data scientist	DNN, transfer learning	Exp.	Q from other lines and new line	Q	Retraining test size divided by 50, training time divided by 4
[56]	Line	Green, brown	Data scientist	DNN	Exp.	Per-channel per-EDFA signal and noise powers (train) then per-channel launch power	NF, OSNR	0.5 dB typ. OSNR error (1.2 dB max.)

Table 3 Comparison of ML-QoT estimation frameworks for networks: classification

Paper	Use Case	Data Source	Learning Type	Algorithm	Type of Study	Inputs (Features)	Output	Key Results
[57]	Network	Brown	Data scientist	SVM	Sim.	Light path length, link lengths, wavelength, statistics on co-propagating light paths	BER (2 classes)	> 99.9% correct decisions
[58]	Network	Brown	Data scientist	ANN	Sim.	Path length, number of EDFAs, max. link length, destination node degree, wavelength	BER (2 classes)	Accuracy > 90%
[39]	Network	Brown	Data scientist	KNN, Random Forest	Sim.	Total length, number of links, maximum link length, demand capacity, modulation format, guardband, and traffic volume of nearest left and right neighbor	BER (2 classes)	Typical accuracy > 95%
[59]	Network	Brown	Data scientist	GPR, active learning	Sim.	Total length, number of links, maximum link length, demand capacity, modulation format	BER (2 classes)	Accuracy > 99% by complementing a small data set of existing services with only a few well-chosen light paths used as probes
[60]	Network	Brown	Data scientist	SVM, ANN	Sim.	Total link length, span length, channel launch power, modulation format, data rate	SNR (2 classes)	> 99% correct decisions with ANN and SVM, ANN 10x faster (sub-ms)
[61]	Network	Brown	Data scientist	KNN, logistic regression, SVM (classification), ANN (classification, regression)	Sim.	Number of hops and spans, light path length, average and maximum link length, average span attenuation, average dispersion, modulation format	SNR (continuous or 2 classes)	Typical classification accuracy > 95%, average SNR error < 0.4 dB
[63]	Network	Brown	Data scientist	SVM	Sim.	Number of ROADMs, of links and of fiber spans, length of fiber span, launch channel power, EDFA pre- and post-amplifier gain, wavelength	OSNR (4 classes)	> 95% correct decisions
[62]	Network	See Section 3.3	Data scientist	SVM, transfer learning	Sim.	Total length, number of links, maximum link length, demand capacity, modulation format	BER (2 classes)	Retraining test size divided by 2

Table 4 Comparison of ML-QoT estimation frameworks for networks: regression

Paper	Use Case	Data Source	Learning Type	Algorithm	Type of Study	Inputs (Features)	Output	Key Results
[68]	Network	Brown	Data scientist	ANN	Exp.	Not explicit	OSNR	Microsecond estimation, RMSE < 0.2 dB in 90% of cases
[29]	Network	Brown	Physicist	Filtering	Exp.	Pre-FEC BER	Q	0.6 dB max. estimation error on Q, millisecond estimation
[69]	Network	Brown	Physicist	Gradient descent	Sim.	Pre-FEC BER, fiber input power, NF	SNR	Typical error < 0.1 dB
[70]	Network	Brown	Data scientist	DNN	Exp.	Channel, noise power on each link	OSNR	< 5 dB OSNR error over range 20-40 dB
[71]	Network	Brown	Physicist; data scientist	Least squares	Sim.	Per span: fiber attenuation, dispersion, nonlinear coefficient (physicist); Number of EDFAs and hops, total path length, baud rate, "load metric" (data scientist)	SNR	SNR RMSE down to < 0.1 dB
[72]	Network	Brown	Hybrid	ANN	Sim.	Gaussian Noise model output, span number, maximum and average lengths, launch power, link length, residual chromatic dispersion, average fiber nonlinear coefficient and attenuation, number of channels	Nonlinear SNR	Error < 2 dB in 95% of the cases
[73]	Network	Brown	Hybrid	ANN	Sim.	Fiber attenuation, dispersion coefficient, effective area, and non-linear refractive index; span length, number of active channels, launch power, channel bandwidth and frequency, output of analytic modeling tool	Nonlinear SNR	> 99.9% estimates within 0.5 dB, < 10 ms computation
[74]	Network	Brown	Data scientist	ANN	Sim.	List of span lengths	SNR	RMSE < 0.2 dB
[75]	Network	Brown	Data scientist	DNN	Exp.	Launch power, laser bias, EDFAs input/output powers	Q	RMSE < 0.02 dB with 600 training points
[76]	Network	Brown	Physicist	SVM regression	Sim.	Power per channel at each node, fine (sub-GHz) optical spectrum every few spans, pre-FEC BER	SNR	Typical error < 0.2 dB
[77]	Network	Brown	Physicist	Gradient descent	Sim.	Pre-FEC BER, spectrum at gain equalizer	SNR	Typical error < 0.1 dB
[78]	Network	Both	Hybrid	ANN	Sim.	Per-channel power and frequency; number of spans; from physical model: ASE and nonlinear noises	SNR	Max. error < 0.5 dB
[79]	Network	See Section 3.3	Data scientist	DNN, evolutionary transfer learning	Exp.	OSA power profile at each WSS	Q	Retraining test size divided by 10
[80]	Network	See Section 3.3	Data scientist	ANN	Sim.	List of span lengths	SNR	RMSE in dB divided by 2
[81]	Network	Brown	Hybrid	DNN	Sim.	Signal powers, channel frequency, number of spans, linear and nonlinear noises (from physical model)	SNR	SNR margin reduction
[82]	Network	Brown	Data scientist	Random forest	Sim.	Mod. Format, traffic volume, light path length, longest link length, number of links	SNR	Estimation of the prob. distribution function of SNR (4 moments)
[83, 84]	Network	Brown	Physicist	Gradient descent	Exp.	Pre-FEC BER, spectrum at each WSS, total EDFA power	SNR	Typical error < 0.2 dB



S+C+L-Band Ultra-high Capacity Optical Transmission

Qiang Guo, Lin Gan, Bo Feng, Shiyi Cao, Xinhua Xiao

B&P Laboratory

Abstract

Extending the fiber spectrum is a promising approach for capacity increasement. Currently, S-band of the optical fiber has attracted a lot of attention from both researchers and practitioners. In this paper, we review the challenges and key technologies of S+C+L optical transmission, including system modeling, S-band optical amplifier, fiber nonlinearity compensation, and flex-rate. In addition, we construct a 150-nm S+C+L-band optical transmission system and evaluate its performance over 1600-km G.652.D and G.654.E fibers through simulation. Then, we experimentally demonstrate 150-nm S+C+L-band optical loop transmission, approaching a fiber capacity higher than 100 Tb/s over 1000 km.

Keywords

S+C+L-band, stimulated Raman scattering, thulium-doped fiber amplifier, bismuth-doped fiber amplifier, flex-rate

1 Introduction

With the deployment of 5G and hyper-scale data centers, the global network traffic is experiencing explosive growth. The fiber capacity needs to be significantly increased to meet the requirements of the growth in network traffic. Expanding the effective transmission spectrum is an optimal way to increase fiber capacity. In addition to the traditional C-band and L-band, S-band or even E-band can be fully utilized to achieve the target of 100-Tb/s capacity for long-haul optical transmission.

In recent years, a lot of academic and industrial institutions are focusing on ultra-wideband optical transmission. Various new optical amplifiers are proposed, such as S-band thulium-doped fiber amplifiers (TDFA) and bismuth-doped fiber amplifiers (BDFA) [1–3], ultra-wideband semiconductor optical amplifiers (SOA), and Raman amplifiers [4, 5]. Several ultra-wideband optical transmission experiments have also been conducted as shown in Figure 1 [6–14]. The system bandwidth and capacity are noticeably in inverse proportion to the transmission distance. It is quite challenging to achieve 100-Tb/s capacity over 1000+ km. The main limiting factors are the optical amplifier noise and fiber nonlinear effects, especially the inter-channel stimulated Raman scattering (ISRS). Wider signal bandwidth and higher signal power will lead to more severe system performance deterioration induced by ISRS. However, shortening the span length of the fiber link (e.g., from 80 km per span to 40–50 km per span) or using G.654.E fibers can significantly decrease the input optical power of the signals. As a result, the ISRS effect can be greatly diminished and the capacity is expected to break the 100-Tb/s limit in long-haul optical transmission applications.

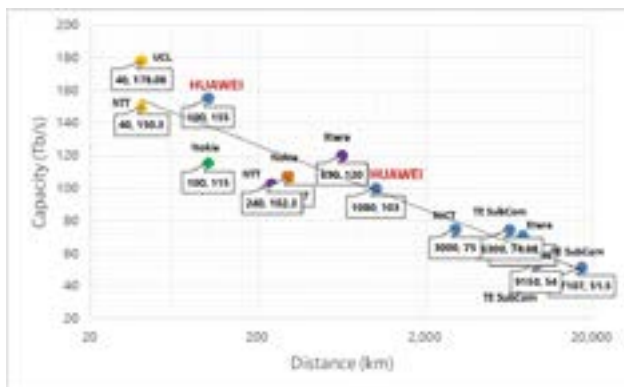


Figure 1 Notable single mode fiber (SMF) ultra-wideband optical transmission demonstrations

In this paper, first we review the challenges and key technologies of S+C+L optical transmission, including ultra-wideband optical transmission system modeling, S-band optical amplifier, fiber nonlinearity issue, especially stimulated Raman scattering, and flex-rate technology for capacity maximization. We have built a novel model based on the traditional Gaussian noise (GN) model. The stimulated Raman scattering (SRS) effect in the optical fiber is taken into consideration to accurately estimate the performance of each wavelength channel in an ultra-wideband optical system. The S-band optical amplifiers are also introduced in this paper, including the TDFA and BDFA. SRS compensation and flex-rate are two key technologies that can optimize ultra-wideband system performance. The former can equalize the performance of all the channels and the latter can maximize the total fiber capacity.

Then we systematically evaluate the performance of S+C+L-band transmission over 1600-km with or without counter-propagating distributed Raman amplification (DRA) in terms of different fiber types and span lengths. We demonstrate that by using the S+C+L with G.654.E fiber of 50-km span length, the total capacity can reach up to 159.6 Tb/s (3.6 times the capacity of C band alone) even without DRA. We also experimentally demonstrate 150-nm ultra-wideband optical recirculating transmission, approaching a fiber capacity of 103 Tb/s over 1080 km. The signal wavelength covers S-band, C-band, and L-band (around 150 nm), and the system consists of 474×32 GbD polarization-division-multiplexed (PDM) probabilistically shaped 16-quadrature-amplitude modulation (QAM) signals (around 169 Gb/s per S-band channel and 205-Gb/s per C or L-band channel). No distributed Raman amplifiers were used in our experiments.

2 Challenges and Key Technologies

2.1 Inter-Channel Stimulated Raman Scattering Gaussian Noise (ISRS GN) Model

The traditional Gaussian noise (GN) model is used for precisely estimating the nonlinear transmission impairments when exploiting state-of-the-art transceivers. However the impact of inter-channel stimulated Raman scattering (ISRS) on the optical Kerr nonlinearity is not taken into

consideration in this model. For an ultra-wideband optical transmission system, such as a C+L or an S+C+L system, the ISRS effect in the fiber is quite obvious. Thus, the traditional GN model is no longer accurate. A novel model named ISRS GN is proposed [15], which accounts for any arbitrary frequency dependent signal power profile. This enables the modeling of nonlinear interference in ultra-wideband regimes where ISRS is significant. The proposed ISRS GN model is a semi-analytical one. It relies on a numerical ordinary differential equation (ODE) solver which is used for calculating the signal power profile.

In this paper, the ISRS GN model is applied to an S+C+L transmission system that occupies the 18 THz optical bandwidth. Figure 2a shows the power transfer of a C+L system and an S+C+L system. As the transmission bandwidth of the C+L system is less than the 14-THz Raman gain spectrum width, the SRS effect causes a tilt of the power spectrum of the wideband optical signals (shown by the blue curve in Figure 2a). The power ripple may reach 5–6 dB. However, the transmission bandwidth of an S+C+L transmission is larger than 14 THz, so the power transfer due to SRS is quite different from that in C+L system. The power ripple even reaches above 10 dB. In addition, as shown in Figure 2b, if the fiber input power is decreased, the SRS-induced power transfer between lower and higher wavelength channels will be suppressed. The simulation results demonstrate that the power transfer due to the SRS is proportional to both the total power and the whole spectral width of the wideband optical signals.

The S+C+L ultra-wideband transmission system suffers from both Kerr nonlinearity and SRS. The interplay between Kerr nonlinearity and SRS is simulated. Figure 3a gives the nonlinear signal-to-noise ratio (SNR) of each wavelength channel with the same fiber input power, without considering the SRS impact. The nonlinearity is relatively weak at longer wavelengths because of the large dispersion. When the SRS effect is included, the nonlinearity at longer wavelengths is enhanced because of the power transfer caused by SRS. The corresponding results are shown in Figure 3b. The proposed ISRS GN model is an efficient tool in ultra-wideband transmission in point-to-point and network scenarios.

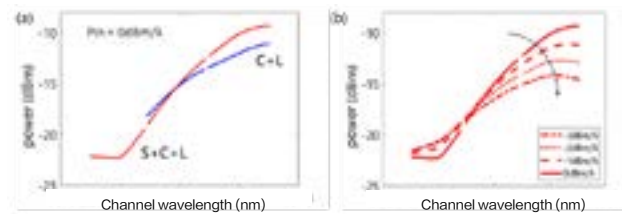


Figure 2 (a) Power transfer due to the SRS effect in S+C+L and C+L systems; (b) Power transfer with different input power in S+C+L system

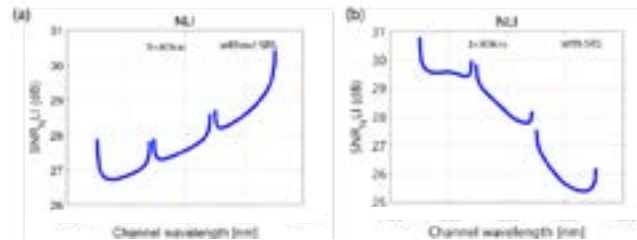


Figure 3 NLI estimation (a) with and (b) without considering the SRS effect

2.2 S-Band Optical Amplifier

S-band is a new low-loss transmission window in addition to the C- and L-bands. Currently, TDFAs and BDFAs are studied for S-band amplification. However, they are not as mature as EDFAs, and new breakthroughs need to be made.

The S-band TDFAs were first reported in 2000 [16, 17]. As the energy gap between the upper energy level and the second-lowest energy level of the thulium (Tm^{3+}) ion is small, the multiphonon relaxation becomes the main process. This makes it difficult to observe the light emitted by Tm^{3+} ions in glass with high phonon energy, such as silicate, borate, and phosphate glass. Therefore, the glasses with low phonon energy, such as fluorozirconate glass, are used as the host material of the doped fibers of S-band amplifiers. In addition, by increasing the Tm^{3+} concentration (e.g., higher than 6000 ppm), the cross relaxation between energy levels of the Tm^{3+} ions is enhanced and the efficiency of the fiber amplifiers can be improved. There are two pumping schemes for the TDFA, namely a 1420-nm pump laser or two pump lasers at wavelengths of 1420 nm and 1560 nm. In addition, as the quartz glass has exceptional physical properties and can be easily processed and spliced with standard transmission fiber, a few groups are still devoted to developing quartz glass-based TDFAs.

BDFA is a promising technology that supports the O-, E-, and S-band amplification and has attracted extensive

attention in recent years. With different host materials, such as aluminosilicate, phosphate silicate, and germanium silicate glass, the emission spectra can cover the 1150 nm to 1500 nm region. O-band and E-band BDFAs have been reported with a spectral bandwidth of 115 nm, a small signal gain of 31 dB, and an NF of 4.8 dB [1]. Y. Wang, D. J. Richardson, et al in [2] demonstrated a germanium silicate glass-based BDFFA, which operates in a spectral range of 1370–1490 nm, covering the E-band and part of the S-band, with a maximum gain of 31 dB and a minimum NF of 4.75 dB. In 2020, an O-band BDFFA was reported. This single-stage, MSA-compliant BDFFA prototype provides up to 18 dB gain and 15 dBm output power over 40 nm gain bandwidth which can be centered between 1290 and 1310 nm [18]. Bismuth-doped silicon dioxide fiber used in BDFFA is similar to the traditional erbium-doped fiber in terms of fiber splicing and mechanical properties. BDFFAs have a simple structure and can be used in many fields, such as telecommunication, data communication, and measurement. Although BDFFAs have many advantages in O-, E-, and S-band transmission, the mechanisms of bismuth ions are not well understood and there is no mature fabrication process available. The commercialization of BDFFA still requires urgent research and development efforts.

Figure 4a and Figure 4b show the performance of a TDFA and a BDFFA built by us, respectively. Currently, the TDFA is facing more engineering problems, such as the fusion splicing of fluoride fiber to silica fiber and the mechanical properties of fluoride fibers. In terms of BDFFA, the mechanisms and reproducibility of BDF should be further studied. If the current problems of the BDFFA can be solved, there is a lot of promise for commercialization of S-band amplifiers.

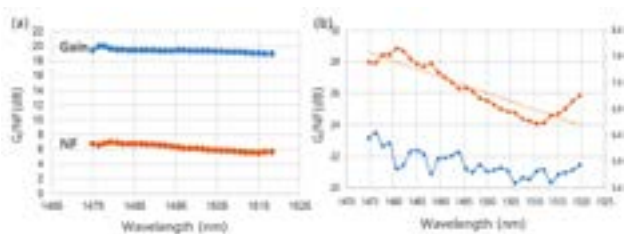


Figure 4 Gain and noise figure of (a) TDFA and (b) BDFFA

2.3 Stimulated Raman Scattering Compensation

The SRS in the optical fiber is an obvious effect for ultra-wideband transmission systems. It leads to significant OSNR degradation of short-wavelength channels and large power unflatness across the entire spectrum [19]. To solve this problem, the fiber input power for ultra-wideband transmission systems should be optimized considering the signal power transition caused by the inter-channel SRS. Thesis [20] gives a scheme of optical power optimization as shown in Figure 5. First, the average fiber input power of the S-band is optimized for single-band transmission by sweeping the power. The power of the C-band is also optimized for single-band transmission because the C-band is almost entirely unaffected by inter-channel SRS for the S- and L-bands. Note that if the optimum power of the S-band changes due to the inter-channel SRS in the final step, the power optimization of the second and the final step is repeated until the optimum powers of the S- and L-bands converge.

In the traditional optical network, reconfigurable optical add/drop multiplexers (ROADMs) exist. Abrupt power fluctuations in the existing channels are inevitable because the asymptotic response of conventional electrically-controlled equalization schemes cannot follow the power fluctuation due to optical nonlinear phenomenon, which may result in instantaneous service interruption. In [21], a technique is proposed to cancel inter-channel SRS by employing wavelength-band inversion. It offsets the power transition created by inter-channel SRS by reversing the direction of power transition at an intermediate site. Besides, another method called "dummy light (DL)" is used to stabilize amplifier behaviors and the SRS effect by presenting to each fiber an unvarying total channel input power spectrum. Every ROADM node is equipped with DL sources (either CW lasers or ASE). WSS filtering on each direction then ensures a constant "full-fill" combination of signal and dummy channels in every optical section. This approach supports system stabilization and fast optical reconfiguration, limited by WSS switching intervals (swapping signal and dummy sources).

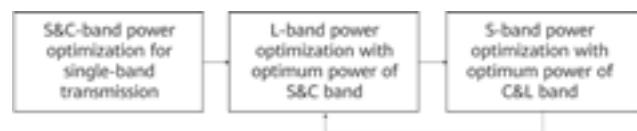


Figure 5 The scheme of optical power optimization in S+C+L system

2.4 Flex-rate Configuration

Flex-rate is quite an important technique for fiber capacity maximization in ultra-wideband transmission systems. Due to the SRS effect, the SNRs of all the channels are wavelength-dependent in long-haul scenarios, as shown in Figure 6. The SNRs of short-wavelength channels are much worse than those of long-wavelength channels because of the SRS-induced power transition and the high NF of S-band amplifiers. In a traditional optical transmission system, all the channels are usually configured with the same modulation format. However, this is not applicable to ultra-wideband transmission systems because the system capacity will be limited by the performance of short-wavelength channels.

Flex-rate technology can adjust the data rate of each wavelength channel according to its SNR. More information is transmitted in high-quality channels than in poor channels, fully utilizing the SNR margin of each channel. The Flex-rate scheme maximizes the total fiber capacity and the transmission distance for ultra-wideband systems.

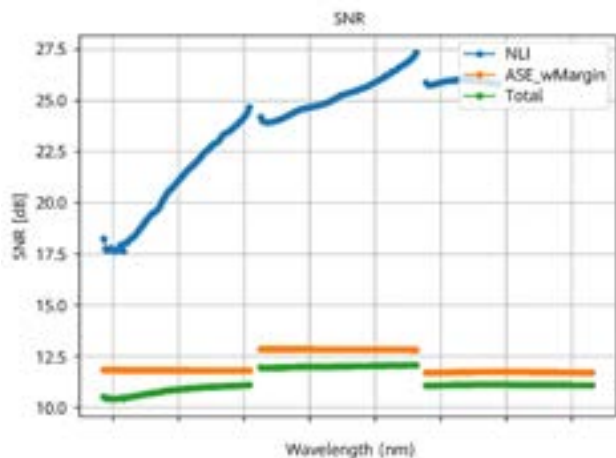


Figure 6 SNRs of all the channels in S+C+L system for long-haul transmission (blue curve for the nonlinearity-induced SNR, orange curve for the ASE-induced SNR, and green curve for the total SNR)

3 Simulation and Experiment

3.1 Simulation

Figure 7 illustrates a simulated S+C+L-band optical transmission system including S-, C-, and L-band transmitters, receivers, re-configurable optical add/drop multiplexers and inline optical amplifiers, and backward

distributed Raman amplifiers (DRAs). The total capacity of each band is determined by its worst channel. When there is no DRA in the system, the mean input power and the power tilt of each band are optimized. When DRA exists in the system, in addition to the mean optical power and power tilt, the Raman pump wavelengths and pump power should be also taken into consideration. The optimization process is performed by the particle swarm optimization (PSO) method. Then the gradient descent (GD) method is used to optimize system capacity and OSNR ripple.

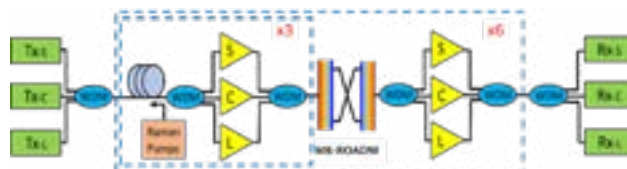


Figure 7 S+C+L-band optical fiber transmission system (Tx: transmitter, WDM: MUX/DEMUX filter, Rx: receiver)

Table 1 shows the performance evaluation of the S+C+L-band transmission over 1600-km G.652.D and G.654.E fiber links. Each band has 80 channels with a channel spacing of 75 GHz. The guard band is 75 GHz between adjacent bands. When G.652.D fiber of 80-km span length is used, the minimum SNR of S-band signals is only 6.1 dB, which is much worse than that of C-band and L-band. In order to improve the performance of S-band signals, G.654.E fiber of 80-km span length can be employed. In this case, the minimum SNR of S-band is increased to 10.3 dB. The similar performance is observed when using G.652.D fiber of 50-km span length. However, the minimum SNR of S-band still does not reach 10.8 dB (the SNR benchmark for C-band transmission). Therefore, we can further improve the S-band transmission performance using G.654.E fiber of 50-km span length. The minimum SNR of S-band is 12.3 dB which is 1.5 dB higher than the benchmark, with a total capacity of 159.6 Tb/s.

Table 1 Minimum SNR and total capacity of each band w/wo DRA

Span Length	Fiber Type	Minimum SNR (dB)			Capacity (Tb/s)		
		S	C	L	S	C	L
80 km	G652.D	6.1	9.8	9.9	27.7	40.2	41.0
	G654.E	10.3	10.9	10.9	42.2	44.4	44.5
50 km	G652.D	10.1	11.2	11.7	41.3	45.7	47.4
	G654.E	12.3	13.8	13.7	49.4	55.2	55.0

3.2 Experimental Demonstration

Figure 8 shows the experimental setup of the ultra-wideband optical transmission system, comprising three parts: the ultra-wideband transmitter, the optical recirculating loop and the corresponding receiver.

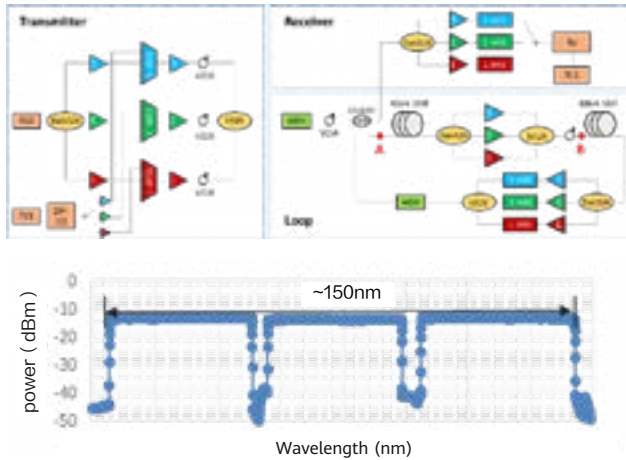


Figure 8 Experimental set-up of 150-nm ultra-wideband recirculating transmission system (top) and the generated 150-nm wideband signal (bottom)

At the transmitter, a high-power and broadband amplified spontaneous emission (ASE) light source is used for generating 37.5-GHz spaced carriers over the S-, C-, and L-bands, as shown in Figure 8. A narrow linewidth (<100 kHz) tunable laser source (TLS) with a 155-nm-wide wavelength tuning range (1475 nm-1630 nm) is used for generating a test channel. Then the test channel is modulated in a dual-polarization in-phase-and-quadrature (DP-IQ) modulator driven by a four-channel arbitrary waveform generator (AWG) operating at 120 GSa/s. A 32-GBd, PDM-16QAM root-raised cosine shaped signal with a roll-off of 0.1 is produced. An S-band waveshaper flattens the spectrum of the S-band dummy channels and carves a notch to accommodate the generated test channel. The C-band and L-band waveshapers do the same. Then the S-band signals are amplified by a TDFA with 24-dBm output power and 7-dB noise figure (NF). The C-band signals are amplified by an EDFA with 22.5-dBm output power and 4.5-dB NF. The L-band signals are amplified by an EDFA with 21.5-dBm output power and a 6.5-dB NF. The amplified S-, C-, and L-band signals are combined together using an S+C+L-band multiplexer and then launched into a recirculating loop.

At the recirculating loop, two acousto-optic modulators (AOMs) are used as optical switches for gating input and recirculating signals. After the first AOM, a 20/80 optical

coupler splits signals for recirculation and reception. The recirculating path contains two fiber spans. The first span length is 40 km and the second one is 80 km. Following each fiber span, an amplification stage is deployed for optical loss compensation. The total launch power and the power tilt of the S-band WDM signal is larger than that of the C- and L-band signal for SRS suppression. After 120 km of fiber transmission, S-, C-, and L-band waveshapers are used for gain equalization. The second amplification stage in the recirculating loop is also used for compensating the loss induced by other optical components, including the waveshapers, the S+C+L-band (de)multiplexers and the second AOM switch.

At the receiver, S/C/L-band signals are separated and then amplified. The test channel is filtered using an optical band-pass filter and detected through coherent detection. The signals after 9-loop (1080 km) transmission were acquired by an 80-GSa/s real-time oscilloscope and then processed offline. In the system, FEC decoding is implemented using a 25% overhead, and the threshold of pre-FEC BER is 3.4×10^{-2} . The SNR of each channel and the system throughput are calculated.

Figure 9 shows the transmission performance of all the channels at a distance of 1080 km, including the information rate and SNR. The mean received SNRs of S-, C-, and L-band signals are 11.3, 13.5 and 14.1 dB, with the measured values in the range 9.5–15.2 dB over the entire transmission bandwidth. S-band signals are configured with probabilistically shaped 16QAM format (SE=3.3) while C- and L-band signals are configured with square 16QAM format (SE=4.0). The aggregate system throughput after decoding is nearly 91.2 Tb/s and the generalized mutual information (GMI) estimated from the acquired SNRs can reach 103 Tb/s.

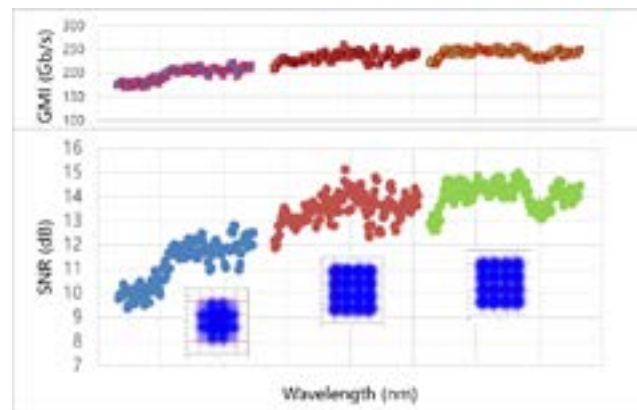


Figure 9 SNRs and data rates of 474 received channels

4 Conclusion

In this paper, we have reviewed the key technologies of S+C+L ultra-wideband optical transmission. The S-band optical amplifier and the ISRS suppression are the major challenges that need to be addressed. We also demonstrate that the total capacity of S+C+L-band transmission can reach more than 3 times that of only C-band transmission when deploying G.654.E fiber with a short span of 50 km. This should be a cost-effective approach to realize multi-band, large-capacity, and long-haul transmission. In addition, we have demonstrated 150-nm ultra-wideband optical recirculating transmission, achieving a fiber capacity of 103 Tb/s over 1080 km for the first time. The results show the potential of wideband optical signals to enable 100+ Tb/s optical transmission over deployed standard single mode fibers in long-haul scenarios.

References

- [1] Yu Wang *et al.*, "Broadband bismuth-doped fiber amplifier with a record 115-nm bandwidth in the O and E bands," in Th4B.1, OFC 2020.
- [2] Y. Wang *et al.*, "Broadband bismuth-doped fiber amplifier with record 117-nm bandwidth in the O- and E-band," in Th4.B1, OFC 2020.
- [3] A. Khagai *et al.*, "O-band bismuth-doped fiber amplifier with 67-nm bandwidth," in W1C.4, OFC 2020.
- [4] J. Renaudier *et al.*, "First 100-nm continuous-band WDM transmission system with 115-Tbps transport over 100-km using novel ultra-wideband semiconductor optical amplifiers," in ECOC 2017, pp. 1-3, 2017.
- [5] J. Renaudier, "Recent advances in ultra-wideband WDM transmission based on semiconductor optical amplifiers," in Tu3F.5, OFC 2019.
- [6] B. J. Puttnam *et al.*, "319 Tb/s Transmission over 3001 km with S, C and L band signals over > 120nm bandwidth in 125 μ m wide 4-core fiber," in F3B.3, OFC2021.
- [7] F. Hamaoka *et al.*, "150.3-Tbps ultra-wideband (S, C, and L bands) single mode fibre transmission over 40-km using >519 Gbps/ λ PDM-128QAM signals," in Mo4G.1, ECOC 2018.
- [8] L. Galdino *et al.*, "Optical Fibre Capacity Optimisation via Continuous Bandwidth Amplification and Geometric Shaping", IEEE PTL, vol. 32, no. 17, pp. 1021-24, 2020.
- [9] M. Ionescu *et al.*, "91 nm C+ L hybrid distributed Raman-erbium-doped fibre amplifier for high capacity subsea transmission," in Mo4G.2, ECOC 2018.
- [10] J. Renaudier *et al.*, "107 Tb/s Transmission of 103-nm Bandwidth over 3 \times 100 km SSMF using Ultra-Wideband Hybrid Raman/SOA Repeaters", in Tu3F.2.1, OFC 2019.
- [11] Akihide Sano *et al.*, "102.3-Tbps (224 \times 548-Gbps) C-and extended L-band all-Raman

- transmission over 240 km using PDM-64QAM single carrier FDM with digital pilot tone," in PDP5C.3, OFC 2012.
- [12] J.-X. Cai *et al.*, "94.9 Tb/s Single Mode Capacity Demonstration over 1900 km with C+L EDFAs and Coded Modulation," ECOC'18, pp. Mo4G.
- [13] Maria Ionescu *et al.*, "74.38 Tbps transmission over 6300 km single mode fibre enabled by C+L amplification and geometrically shaped PDM-64QAM," *Journal of Lightwave Technology* 38, 531-537, 2019.
- [14] Jin-Xing Cai *et al.*, "70.46 Tbps over 7,600 km and 71.65 Tbps over 6,970 km transmission in C+L band using coded modulation with hybrid constellation shaping and nonlinearity compensation," *Journal of Lightwave Technology* 36, 114-121, 2018.
- [15] D. Semrau, *et al.*, "A Modulation Format Correction Formula for the Gaussian Noise Model in the Presence of Inter-Channel Stimulated Raman Scattering," *J. Lightwave Technol.* 37, 5122-5131, (2019).
- [16] S. Aozasa *et al.*, "Tm-doped fiber amplifiers for 1470-nm band WDM signals," *IEEE Photonics Technology Letters*, Vol. 12, No. 10, 2000.
- [17] T. Sakamoto *et al.*, "High gain and low noise TDFA for 1500 nm band employing novel high concentration doping technique," *Optical Fiber Communication Conference*, Optical Society of America, 2000.
- [18] Vitaly Mikhailov *et al.*, "Simple broadband bismuth doped fiber amplifier (BDFA) to extend O-band transmission reach and capacity," *Optical Fiber Communication Conference*. Optical Society of America, 2019.
- [19] H. D. Kim and C. H. Lee, "Capacities of WDM transmission systems and networks limited by stimulated Raman scattering," *IEEE Photonics Technol. Lett.* 13, 379-381 (2001).
- [20] Hamaoka, Fukutaro, *et al.*, "Ultra-wideband WDM transmission in S-, C-, and L-bands using signal power optimization scheme," *Journal of Lightwave Technology* 37, 1764-1771 (2019).
- [21] Kawahara, Hiroki, *et al.*, "Cancellation of Static and Dynamic Power Transitions induced by inter-band Stimulated Raman Scattering in C+L-band WDM Transmission," *Opto-Electronics and Communications Conference*. IEEE, 2020.



Performance Challenges and Key Technologies of All-Optical Cross-Connect Network Filtering Effect

Ning Deng ¹, Luo Han ¹, Jingquan Xu ¹, Te'an Lu ¹, Di Xiong ¹, Bowen Zhang ¹, Wei Jia ²

¹ Optical Research Dept

² Optical Architecture & Technology Planning Dept

Abstract

The rapid development of services such as 5G, F5G, data center, and cloud interconnection drives the exponential growth of data traffic, which in turn poses great challenges on the cross-connect grooming capabilities of traditional network nodes [1–3]. In recent years, building on the reconfigurable all-optical add/drop multiplexer (ROADM) system, Huawei and industry partners have introduced new technologies such as the wavelength selective switch (WSS) and add/drop wavelength selective switch (ADWSS) that feature more degrees, all-optical backplane, and digital optical layer. All-optical cross-connect (OXC) is becoming the overarching trend of next-generation networks [4–6]. The OXC network uses fiber-free interconnection and colorless, directionless, contentionless, and gridless (CDCG) features [7–9] to satisfy the requirements of ultra-large cross-connect capacity, high integration, many degrees, and intelligent O&M management.

1 OXC Performance Parameters

An OXC network implements ultra-large-capacity data transmission through end-to-end all-optical domain transmission and wavelength switching. A WSS is the core component of a switching node. The key performance parameters that affect signal quality include insertion loss (IL), polarization-dependent loss (PDL) [10], extinction ratio/crosstalk [11], and filter passband characteristics [12, 13], as shown in Figure 1. IL refers to the energy loss of wavelength division multiplexing (WDM) signals caused by a WSS. Such loss is mainly caused by the dispersion unit, liquid crystal on silicon (LCoS) component, fiber coupling, and system assembly errors within the WSS. Polarization-dependent loss refers to the difference in energy transmitted by a WSS component in two orthogonal states of polarization, and is mainly related to polarization division processing in WSS optical design. Crosstalk refers to a portion of energy of a WDM channel that appears on a non-target port when the WSS allocates the WDM channel to a target port, and is mainly affected by the order of LCoS's high-order diffraction. The filter passband characteristics of WSS mainly include bandwidth and frequency offset. Generally, the spectrum width corresponding to the attenuation values of -0.5 dB and -3 dB in the filter passband is defined as the filter passband width (bandwidth), and the offset between the center frequency of the filter passband and the standard wavelength of an optical signal is defined as the frequency offset. The filter passband characteristics are related to the

LCoS size, WSS spectral range, light spot size, and operating environment.

Essentially, the WSS is a tunable filter, and the bandwidth and filter frequency offset of the WSS mainly affect the signals. On a large OXC network, a WDM signal may traverse 15 ROADMs/OXC sites (30 WSSs) before reaching the destination. A slight difference in the filter passband characteristics of a single WSS can affect the channel quality at the network layer. The WSS filter characteristics and center frequency offset of each WSS determine the transmission capacity, transmission distance, and all-optical network scale of an OXC network to a certain extent.

As the filter bandwidth of a single WSS increases, the filter window is enlarged. Under the same frequency offset condition, the filter bandwidth after multiple WSSs are cascaded in an optical communication link is increased, whereas the filter effect on WDM signals is reduced. When the center frequencies of multiple WSS filter windows on a link are offset, the cascaded filter window shrinks. As a result, the bandwidth of the cascaded filter spectrum decreases, in turn causing additional filtering and affecting the signal quality. In addition, as shown in Figure 2, a filter frequency offset of the WSS further causes an offset of a WDM signal in a filter window after cascading, and asymmetric filtering splits a channel spectrum, which further causes performance deterioration in a transmission system. As such, both increasing the WSS filter bandwidth and reducing the WSS filter frequency offset can effectively improve OXC network performance and support a larger all-optical network scale.

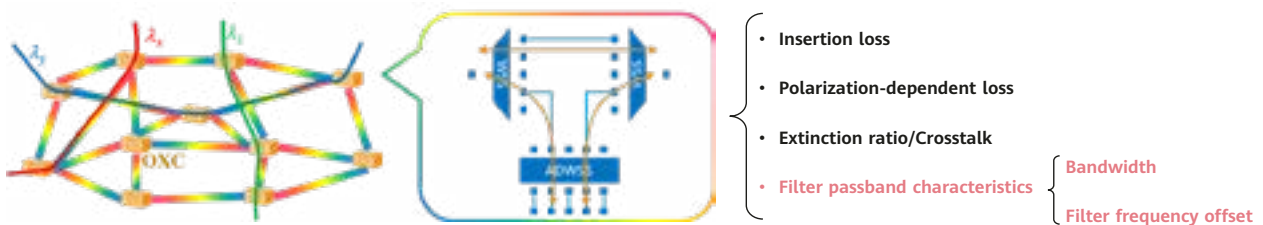


Figure 1 Key performance parameters that affect the signal quality of an OXC network

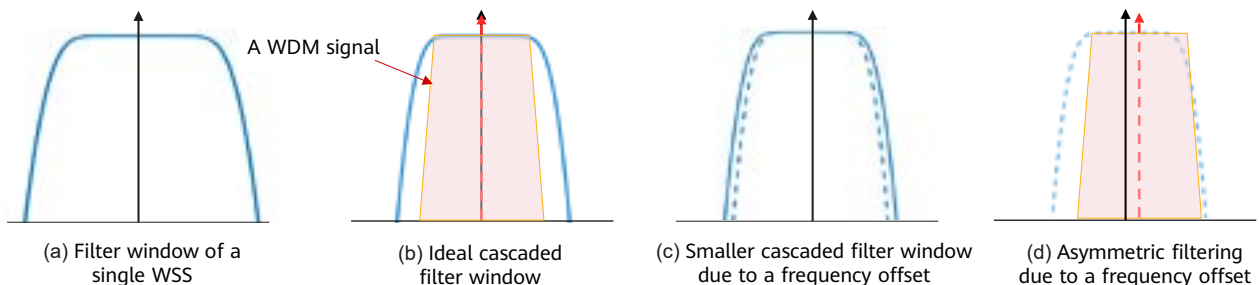


Figure 2 Impact of WSS filter window's frequency offset on WDM signals

2 Innovative OXC Bandwidth Improvement Solution

2.1 Related Solutions in the Industry

OXC hardware consists of optical line boards, optical backplanes, and optical tributary boards. The most important component is the wavelength selective switch (WSS). Currently, a WSS in an OXC mainly uses a liquid crystal on silicon (LCoS) as a switching engine. The LCoS is a pixelated phase-type spatial light modulator (SLM), which features the flexible grid function. In addition, the LCoS implements non-mechanical beam deflection by means of phase modulation, and features high reliability.

Bandwidth is an important indicator for measuring the performance of a WSS. A WSS is considered as a filter, and a frequency response formula $B(f)$ of a WSS can be expressed as [14]:

$$B(f) = R(f) \otimes L(f)$$

In the formula, f is the optical frequency, $L(f)$ is the optical transfer function determined by the light spot size of the LCoS, and $R(f)$ is the filter window function of the LCoS. For ideal passband filtering, $R(f)$ can be expressed as the rectangular function:

$$R(f) = \begin{cases} 1, & -B/2 \leq f \leq B/2 \\ 0, & \text{otherwise} \end{cases}$$

In the formula, B is the frequency width of the filter window. The light field strength distribution of a light spot on the LCoS is usually a Gaussian function, which is expressed as follows:

$$L(f) = \exp\left[-\frac{f^2}{2\sigma^2}\right]$$

In the formula, σ is the waist radius of the light spot. According to the frequency response expression of WSS filtering, the bandwidth performance of OXC is mainly determined by the light spot size of the LCoS and the filter window function.

Currently, there are two technical paths in the industry for improving OXC bandwidth performance, as shown in Figure 3. One technical path is to compress the size of a light spot on the LCoS by means of optical design. Compressing the light spot size on LCoS increases the divergence angle of the light spot in the dispersion direction, decreases the optics assembly tolerance, and increases the assembly complexity. Therefore, compressing the light spot size through optical design offers only limited bandwidth improvement.

The other technical path is to use the LCoS algorithm to improve the bandwidth. The WSS can be considered as an optical spectrum processor (OSP). It can separately adjust the attenuation of the spectrum signal on each pixel to achieve flexible spectrum shaping. The LCoS algorithm can shape the spectrum to form a filter spectrum with higher bandwidth, thereby improving the equivalent bandwidth, as shown in Figure 4. However, in achieving bandwidth improvement, this method increases the insertion loss of the system. But considering that an optical amplifier (OA) in a system can compensate for the insertion loss of a link, the LCoS algorithm is more effective in improving bandwidth.

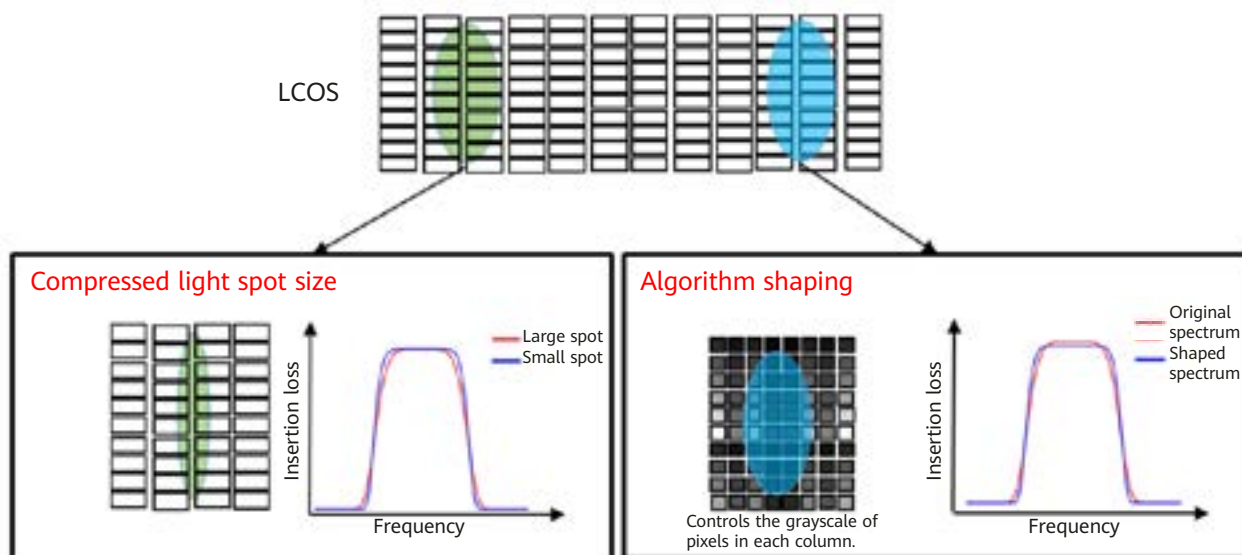


Figure 3 Technical paths to improve OXC bandwidth performance

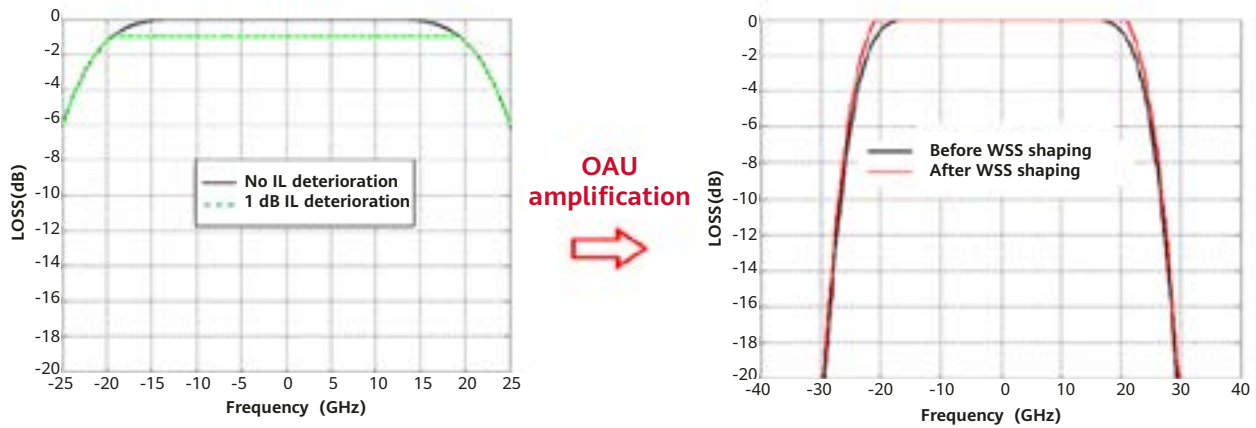


Figure 4 Basic principles of the bandwidth improvement technology based on spectrum shaping

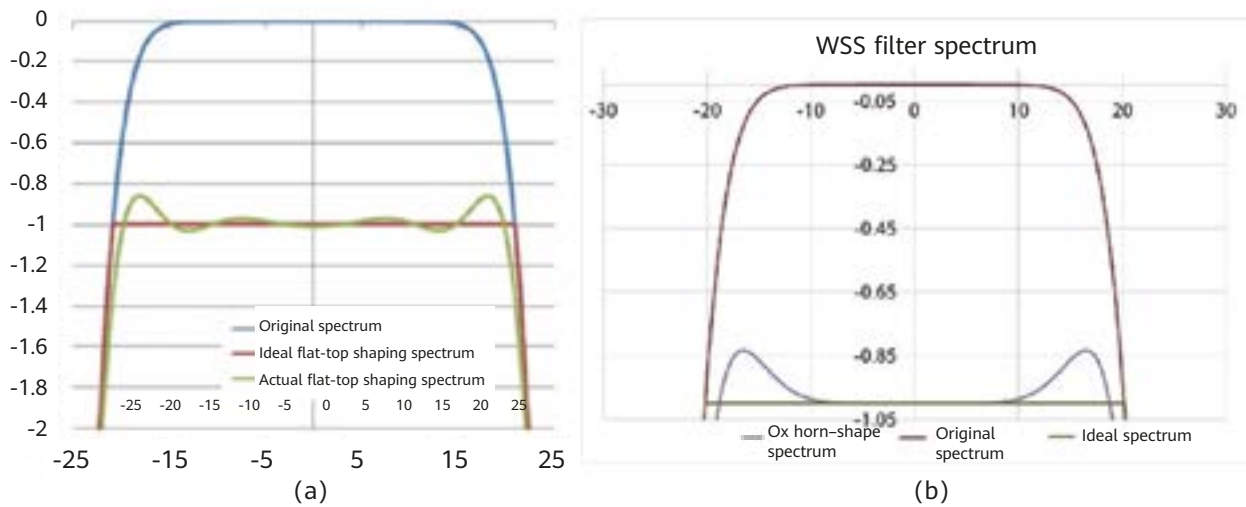


Figure 5 (a) Schematic diagram of flat-top spectrum shaping; (b) Schematic diagram of "ox horn-shape" spectrum shaping

Currently, all commercial WSS modules use bandwidth improvement algorithms based on spectrum shaping [15]. There are two common solutions in the industry. One is a flat-top spectrum shaping algorithm, and the other is an "ox horn-shape" spectrum shaping algorithm. The flat-top spectrum shaping algorithm cuts the spectrum in one-size-fits-all mode and controls the attenuation value of each column of pixels to reduce the insertion loss in the passband to the same level. However, due to the overlapping of light spots in the LCoS, an increase in the insertion loss of a spectral component of one column of pixels affects the insertion loss of spectral components of other columns of pixels, and causes an insertion loss fluctuation (IL ripple) when increasing the bandwidth, as shown in Figure 5 (a). Figure 5(b) shows the schematic diagram of an "ox horn-shape" spectrum shaping algorithm. With this algorithm, a spectrum within an effective bandwidth (clear passband) is flat-top shaped, a spectrum outside the effective bandwidth is not attenuated,

and the finally-obtained spectrum is in an "ox horn" shape. This is why it is referred to as an "ox horn-shape" spectrum shaping algorithm. This shaping algorithm can effectively suppress the insertion loss fluctuation within the bandwidth. However, the out-of-band insertion loss fluctuation (OOB ripple) is large, with insertion loss fluctuation occurring at the "horns".

2.2 Innovative Solution to Increase WSS Bandwidth Through Spectrum Shaping

To address the high insertion loss fluctuation of bandwidth improvement algorithms in the industry, Huawei establishes three evaluation dimensions: least mean square (LMS) of the filter spectrum and ideal spectrum, insertion loss fluctuation, and bandwidth. In this way, Huawei implements

the optimal iterative spectrum shaping, achieving optimized LMS (OLMS) spectrum shaping. First, a spectrum function is optimized based on the LCoS light spot parameters and initial pixel insertion loss. The LMS of the current spectrum function and ideal spectrum function, insertion loss fluctuation, and bandwidth are calculated, and weights of the three parameters are set to generate an evaluation function. Then the evaluation function is fed back to the LCoS to adjust the pixel insertion loss. The current spectrum function is also evaluated to form an iteration, and finally the spectrum is shaped.

Table 1 Performance parameters of different spectrum patterns

	Insertion Loss Fluctuation (dB)	-3 dB Bandwidth (GHz)	-0.5 dB Bandwidth (GHz)
Original spectrum	0	46.675	40.3
Ideal spectrum pattern	0	47.95	43.9
Ox horn-shape spectrum pattern	0.3434	47.45	42.25
OLMS spectrum pattern	0.0336	47.675	42.475

Figure 6 shows the spectrum obtained by using the optimal iterative spectrum shaping algorithm and the spectrum obtained by using other methods. The filter spectrum obtained by using this algorithm is closer to the ideal spectrum, and has lower insertion loss fluctuation than that obtained by using the "ox horn-shape" spectrum shaping algorithm. Table 1 lists the performance parameters of different spectrum types. When the optimal iterative spectrum shaping algorithm is used, the -3 dB bandwidth gain reaches 1 GHz, and the -0.5 dB bandwidth gain reaches 2.175 GHz. Compared with the "ox horn-shape" spectrum shaping algorithm, the optimal iterative spectrum shaping algorithm slightly increases the -3 dB bandwidth, but significantly reduces insertion loss fluctuation from 0.3434 dB to 0.0336 dB. The optimal iterative spectrum shaping algorithm considers three factors: LMS, insertion loss fluctuation, and bandwidth. Different weights can be set for these factors to obtain an adaptive optimized spectrum pattern. The algorithm can select appropriate spectrum patterns for different application scenarios, ensuring the

bandwidth performance of signals. Compared with other solutions in the industry, this solution has higher bandwidth and lower signal filter impairment, and the advantages are more prominent after WSS cascading. For these reasons, this solution has great advantages in OXC networks.

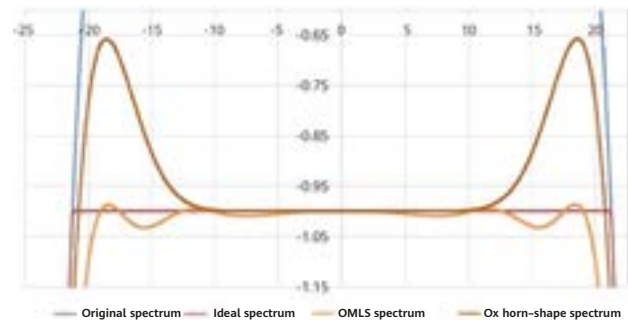


Figure 6 Spectrum patterns of different spectrum shaping algorithms

3 Innovative Solution for Improving OXC Frequency Offset Performance

3.1 WSS Frequency Offset Performance Impact and Status

The frequency offset performance of WSSs is a key technical specification that affects OXC filter passband characteristics. In an ROADM/OXC network, service signals usually travel through multiple nodes before reaching the client side. Therefore, even a slight deterioration of WSS frequency offset performance will cause significant signal quality deterioration after cascading [16]. For a WSS with a grating-lens-LCoS structure, due to the stress and strain caused by factors such as external ambient temperature and atmospheric pressure changes as well as aging of glue inside the WSS, a slight angle, displacement, or refractive index change occurs on components. As shown in Figure 7, a light spot incident on the LCoS moves, causing the WSS filter channel to deviate, which in turn affects the WSS frequency offset performance. For WSS frequency offset correction, a common practice in the industry is to use a temperature or barometric pressure sensor to monitor the internal environment of a WSS in real time. The correction is performed based on the pre-calibrated relationship between the temperature or barometric pressure and the frequency offset. However, such technical methods indirectly detect frequency offsets, and cannot detect the frequency offset caused by various complex factors. In this paper, a high-precision solution for direct frequency offset detection and correction is proposed.

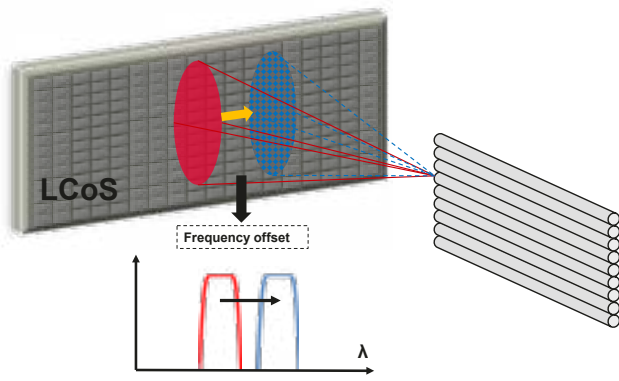


Figure 7 WSS frequency offset generation

3.2 Low-Frequency-Offset WSS Solution Based on Wavelength BeiDou

3.2.1 Frequency Offset Detection and Correction Technology

Theoretical simulation and experimental verification show that WSS frequency offset changes are consistent due to the displacement and expansion of the grating, lens, and LCoS caused by temperature, atmospheric pressure, and aging factors, as well as gas refractive index changes. In Figure 8, the blue line indicates the frequency offsets of different wavelengths at different ports of the WSS in the beginning of life (BOL) state, and the red line indicates the frequency offset changes that are simulated by changing the ambient temperature in the end of life (EOL) state. The frequency offsets of wavelengths at each WSS port are consistent. According to this characteristic of WSS frequency offset changes, on the premise that signals are not affected, the supervisory signals of an out-of-band channel can be used to detect and represent an overall frequency offset change amount of the WSS in real time. To implement frequency

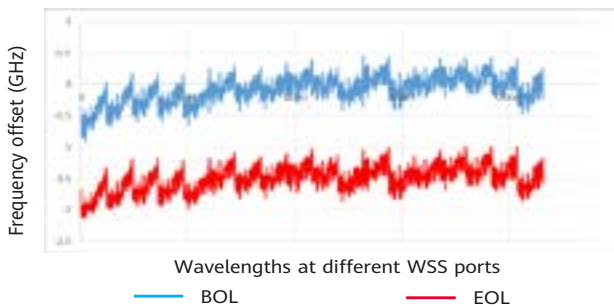


Figure 8 Frequency offset changes of the WSS in the BOL and EOL states

offset correction, the mapping relationship between each wavelength and a pixel area is adjusted based on the change amount by using an LCoS control algorithm. The key to realizing high-precision frequency offset detection is to improve the wavelength accuracy of supervisory signals. This paper proposes a high-precision "wavelength BeiDou" module as the wavelength scale in inspection and detection.

3.2.2 Wavelength BeiDou Technology

The wavelength detection precision of traditional Etalon is limited [17], and the typical value is about ± 2.5 GHz. The factors that limit the precision include the noise power tolerance, Etalon response curve slope, and long-term reliability. The formula is as follows:

$$\Delta\lambda = \frac{\Delta P}{k}$$

In this formula, ΔP is the power resolution, which is mainly affected by noise, and k is the absolute value of the Etalon response curve slope. Currently, the Etalon specifications cannot meet the requirements of WSS frequency offset detection. To further improve the wavelength detection precision, the wavelength BeiDou module proposed in this solution implements high slope compensation of the response curve in continuous wavelengths based on the free spectral range (FSR) response curve's differential effect of the innovative Etalon. A commercial wavelength meter with a wavelength precision of ± 0.3 pm is used to test continuous wavelengths in the C band. The wavelength detection precision can reach ± 2 pm, which is 10 times higher than the wavelength precision of the traditional Etalon. Figure 9 shows the test result.

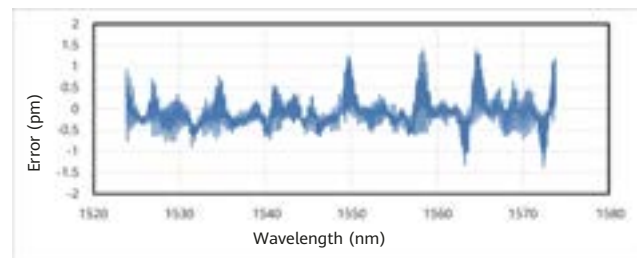


Figure 9 Detection error of wavelength BeiDou in the C band (measured result: $< \pm 1.5$ pm (< 0.19 GHz))

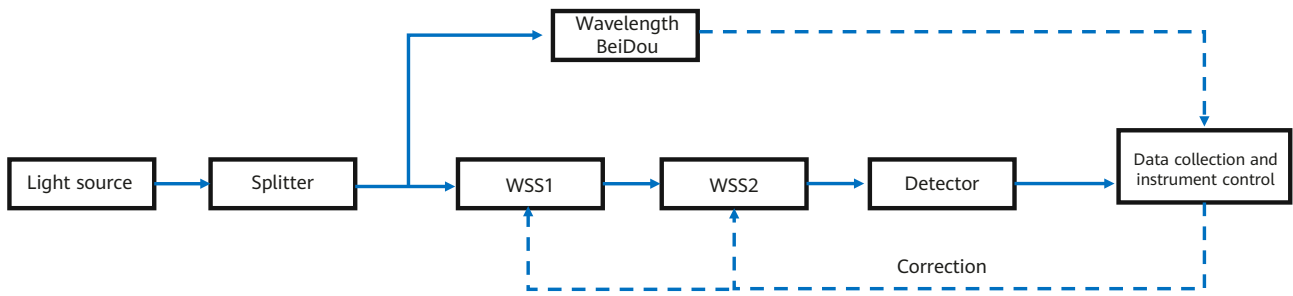


Figure 10 Schematic diagram of an experimental apparatus for a low frequency offset subsystem based on the wavelength BeiDou module

3.2.3 Verification of the Low-Frequency-Offset WSS Subsystem

Figure 10 is the schematic diagram of verifying a low-frequency-offset WSS subsystem based on a wavelength BeiDou module. A light source generates an out-of-band communication supervisory signal, which is split into two channels of signals by an optical splitter. One channel passes through a wavelength BeiDou module, and the other channel passes through two cascaded WSSs and a detector. The two channels of signals are collected and processed by using an algorithm to obtain the out-of-band communication filter spectrum A of the WSS. After a frequency offset occurs in the WSS, the out-of-band communication filter spectrum B of the WSS is obtained through measurement using the same process. The frequency offset change amount of the WSS can be obtained by analyzing the drifts of the out-of-band filter spectrums A and B, and the mapping relationship between a wavelength and an LCoS pixel area is adjusted through closed-loop feedback to correct the frequency offset of the WSS. Figure 11 (a) shows a frequency offset correction result and a correction error based on the case in Figure 8. The green line indicates the frequency offset after the WSS frequency offset change amount is detected and corrected by the wavelength BeiDou module. This line almost overlaps with the blue line (the BOL state), realizing high-precision frequency offset compensation. Figure 11 (b) shows the error curve of frequency offset correction. The correction precision is within ± 0.5 GHz. This means that if a WSS is affected by the frequency offset caused by the external environment and component aging in its lifecycle, the WSS can detect the overall frequency offset change through the wavelength BeiDou module and correct the change using the LCoS algorithm, so that the EOL frequency offset specification of the WSS can be reduced from ± 2.5 GHz to ± 1.5 GHz. This method effectively improves the WSS

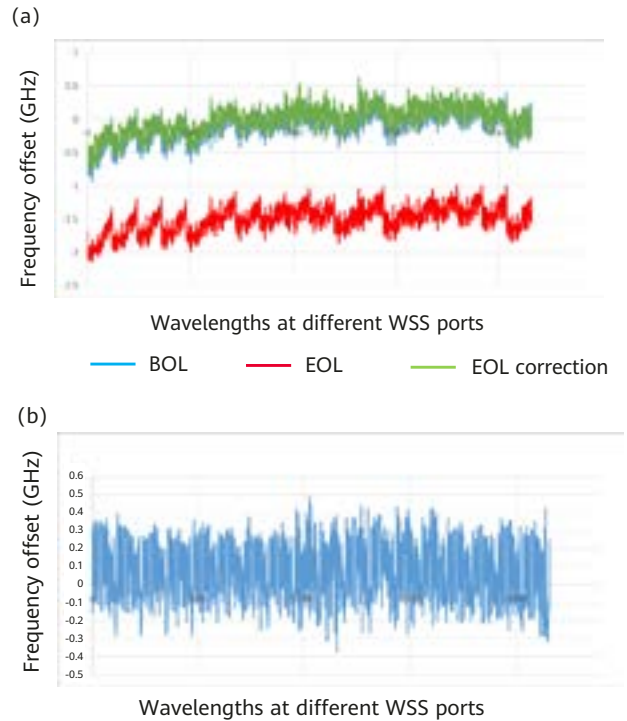


Figure 11 (a) Correction result of the WSS frequency offset in an EOL state; (b) WSS frequency offset correction error

frequency offset specifications, increases the WSS cascading filter spectrum bandwidth, reduces the overall frequency offset, and improves the network transmission performance.

4 Verification of Benefits to a High-Performance OXC System

4.1 Overview of the High-Bandwidth and Low-Frequency-Offset WSS Solution

As a WSS is the core component of a switching node, its filter spectrum performance has a significant impact on a ROADM transmission network. The main characteristics of the WSS filter spectrum include bandwidth and frequency

offset. The spectrum shaping technology can effectively increase the filter spectrum bandwidth of WSSs and reduce the impact of the filter spectrum on the signal spectrum. Its system optical signal-to-noise ratio (OSNR) gain is about 0.6 dB (200G e16QAM). The low-frequency-offset WSS solution based on wavelength BeiDou reduces the WSS filter spectrum frequency offset, increases the filter spectrum bandwidth of WSS cascading, and reduces the overall frequency offset between the cascading filter spectrum and OTU. Its system OSNR gain is about 0.5 dB (200G e16QAM). The combination of spectrum shaping and wavelength BeiDou technologies can optimize both the bandwidth and frequency offset of the filter spectrum of a WSS. In this case, the system OSNR gain is higher than 1 dB (200G e16QAM), effectively improving the transmission link performance and supporting a larger network scale.

4.2 System Benefit Analysis of the WSS Solution with High Bandwidth and Low Frequency Offset

4.2.1 System Benefits of the Spectrum Shaping Technology

In an optical communications network based on WDM technologies, a signal usually traverses a large number of nodes to reach a destination, and even a slight difference in filter bandwidth may have a sizeable impact after cascading. As such, the filter bandwidth plays an important role in the entire optical network. The spectrum shaping technology optimizes the filter spectrum to increase the -3 dB bandwidth gain to 1 GHz and the -0.5 dB bandwidth gain to 2.175 GHz. In addition, this technology guarantees low insertion loss fluctuation, effectively improving the bandwidth performance of the WSS filter spectrum. Here, the cascading of 20 WSSs is analyzed as an example. The cascading filter spectrums before and after bandwidth improvement are compared, and the system transmission characteristics under corresponding filter spectrums are simulated. The result shows that the spectrum shaping technology brings an OSNR gain of about 0.6 dB (200G e16QAM).

4.2.2 System OSNR Gain of the Low-Frequency-Offset WSS Solution Based on the Wavelength BeiDou Technology

A WSS frequency offset affects the overall frequency offset and WSS cascading bandwidth of transmission links. Therefore, the preceding two factors need to be considered when analyzing the impact of a low-frequency-offset WSS solution on the transmission link performance.

(1) Analysis of overall frequency offset benefits

The current OTU frequency offset specification is ± 1.5 GHz, and the WSS frequency offset specification is ± 2.5 GHz. The typical overall frequency offset on live networks is 2 GHz. With the low-frequency-offset WSS solution, the WSS frequency offset specification is reduced from ± 2.5 GHz to ± 1.5 GHz, and the typical overall frequency offset is reduced to about 1.5 GHz.

(2) Analysis of bandwidth benefits

The WSS frequency offset affects the WSS cascading filter spectrum bandwidth. The WSS cascading filter spectrums under different frequency offset specifications are simulated to evaluate the WSS cascading bandwidth gain of the low-frequency-offset solution based on the wavelength BeiDou technology. The frequency offset statistics after WSS aging show that the frequency offset distribution of WSS EOL roughly complies with the super-Gaussian distribution. The WSS cascading bandwidth gain is calculated based on the frequency offset distribution simulation. The simulation solution is as follows:

1. Before and after frequency offset reduction, randomly select the EOL frequency offsets of 20 WSSs from the super-Gaussian distribution according to the corresponding frequency offset distribution, and superpose the filtering spectrums of corresponding frequency offset values to obtain the cascading filter spectrum and its bandwidth.
2. Repeat the preceding process N times to obtain N groups of cascading filtering spectrums before and after frequency offset reduction and N groups of cascading filter spectrum bandwidths respectively.
3. Sort the cascading filter spectrums before and after frequency offset reduction in order of bandwidth.
4. Select the bandwidth value that is lower than or equal

to 99.7% of all bandwidth values from the two groups of bandwidths, use this value to calculate the bandwidth gains of cascading filter spectrums before and after frequency offset reduction, and record the cascading filter spectrums.

(3) Analysis of transmission link OSNR benefits

To calculate the benefits brought by the wavelength BeiDou-based low-frequency-offset WSS solution to transmission links, the bandwidth and frequency offset must be considered. The OSNR penalty curves of two cascading filter spectrums are obtained using the preceding simulation method (simulation condition: 200G e16QAM). Figure 12 shows the simulation result. Considering that the overall frequency offset is reduced from 2 GHz to 1.5 GHz, the system OSNR gain is about 0.5 dB, which matches the test result.

(4) Extreme scenarios

The preceding statistics collection method does not draw on the benefits of the cascading bandwidth in extreme scenarios. The frequency offset specification of the WSS decreases from ± 2.5 GHz to ± 1.5 GHz after the frequency offset reduction. According to the analysis of the WSS cascading filter spectrum bandwidths in two extreme frequency offset scenarios, the -3 dB bandwidths of the WSS cascading filter spectrums are 30.2 GHz and 31.5 GHz, respectively. In this scenario, the filter spectrum bandwidth gain reaches 1.3 GHz, and the system OSNR gain is about 1.2 dB (200G e16QAM).

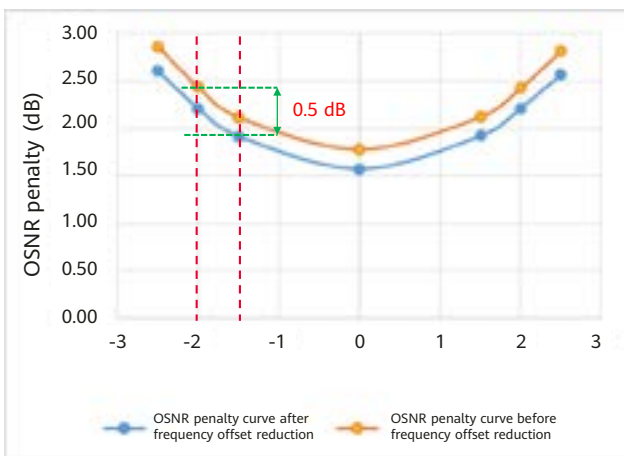


Figure 12 OSNR penalty curves of WSS cascading filter spectrums

5 Future Research Prospects

The optimized filtering effect (such as bandwidth and frequency offset) of WSSs supports high-performance large-granularity service grooming. The signal adjustment function of WSSs enables OXC to implement spectrum signal processing while grooming network wavelengths flexibly. As the transmission rate of signals advances to 400G+, signals are more sensitive to noise and distortion. The requirements for electrical signal processing are constantly increasing, and the digital signal processing (DSP) algorithm of electrical chips grows in complexity. If the WSS-based optical domain control and electrical chip DSP are combined to perform optical electrical signal processing (OESP), the complexity of electrical signal processing can be effectively reduced and the signal transmission quality can be improved, thereby reducing the OSNR penalty. As shown in Figure 13, for 400G+ signals with a high-baud-rate modulation format, the bandwidth limitation of electrical components, hole burning effect of optical fiber amplifiers, coating error of gain flattening filters (GFFs), and stimulated Raman scattering (SRS) effect all cause large ripples in the signals and introduce inter-symbol interference. Utilizing the spectrum adjustment capability of WSSs, frequency domain equalization can be performed for signals at OXC nodes to achieve flat signal transmission [18, 19]. In addition, the DSP algorithm is mainly used to compensate for fiber dispersion, which requires high power consumption. The WSS can control the drive voltage of an LCoS to change phases, implement dynamic dispersion compensation for multiple channels, and fully utilize the computing power of electrical chips [20].

Using the bandwidth limitation of an electrical component as an example, due to the physical limitation of a complementary metal-oxide-semiconductor (CMOS) process, a bottleneck exists between the bandwidth of components such as digital-to-analog converter (DAC), driver, and modulator and the bandwidth required by signals. Consequently, the high-frequency component attenuates severely after signal modulation. 400G+ signals use high-baud-rate modulation formats, and the impact of component bandwidth limitations is also more severe. When the DSP algorithm of electrical chips is used for bandwidth compensation, the peak-to-average power ratio (PAPR) of signals is increased and the quantization distortion of the DAC is introduced. A high-resolution WSS can set a filter spectrum based on the frequency response of a component bandwidth and perform frequency domain equalization on

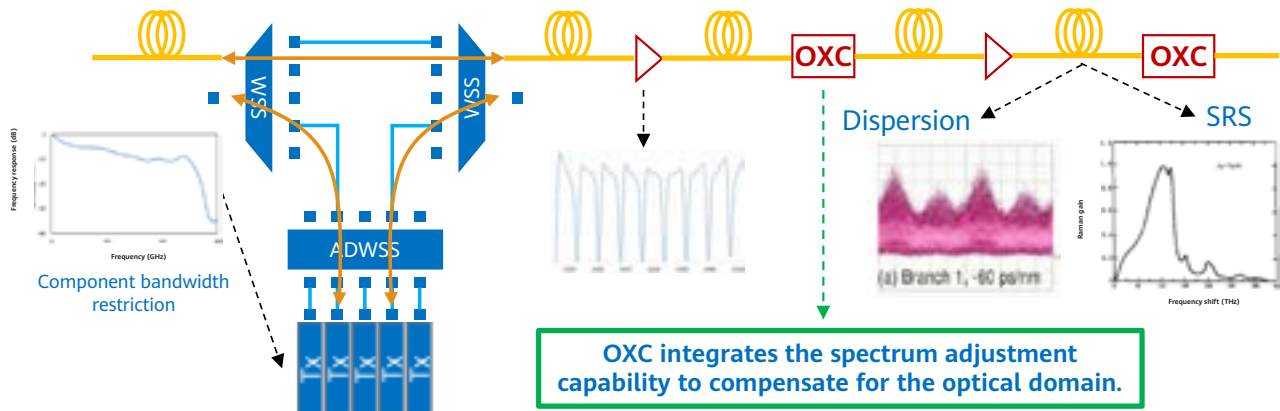


Figure 13 OESP mechanism

a signal spectrum. In this way, the quantization distortion of the DAC can be reduced, a signal-to-noise ratio can be improved, and the algorithm complexity and power consumption requirement of electrical chips can be lowered. Bandwidth compensation based on WSSs has been reported. Compared with the algorithm equalization of electrical chips, this solution achieves 1.7 dB, 2.1 dB, and 4.6 dB OSNR tolerance gains for 64Gb/s PDM-256QAM, 88Gb/s PDM-64QAM, and 92Gb/s PDM-64QAM modulation formats, respectively [18]. In the future, with the in-depth research of OESP, the diversified applications of WSSs will significantly improve the next-generation optical transmission performance and promote the continuous evolution of an "optical Moore's Law".

6 Conclusion

All-optical cross-connect communication networks are advancing towards more degrees and higher rates, in turn posing higher requirements on OXC and its core functional unit WSS. In addition to low insertion loss and many degrees, the filter characteristics (bandwidth and frequency offset) of the OXC or WSS play a decisive role in the OXC network scale and transmission distance. The improvement of WSS filtering effect reduces the filter impairment on WDM signals and effectively improves the scale, performance, and capacity of OXC communication networks. In this paper, the source of the filter effect and system impairment in all-optical cross-connect networks are systematically analyzed, and an innovative high-bandwidth and low-frequency-offset WSS solution is proposed. The spectrum shaping technology increases the WSS filter bandwidth by 1 GHz while ensuring a low insertion loss fluctuation (< 0.035 dB). In the scenario where 20 WSSs are cascaded, the system OSNR gain is about 0.6 dB

(200G e16QAM). Compared with the traditional Etalon, a wavelength BeiDou module improves the wavelength detection precision by 10 times, reaching ± 2 pm. The frequency offset specification of a low-frequency-offset WSS solution based on the wavelength BeiDou technology is reduced from ± 2.5 GHz to ± 1.5 GHz. In the scenario where 20 WSSs are cascaded, the system OSNR gain is about 0.5 dB (200G e16QAM). The combination of spectrum shaping and wavelength BeiDou technologies can further improve the filtering effect of OXC networks. In the future, the WSS's optical domain control capability and OESP will be utilized to significantly improve the performance of ultra-large-capacity optical communication networks, promoting the next-generation WDM technology and industry development later than the generation of single-wavelength 400G (such as 800G and 1.6T).

References

- [1] Gringeri S, Basch B, Shukla V, et al. Flexible architectures for optical transport nodes and networks[J]. IEEE Communications Magazine, 2010, 48(7): 40-50.
- [2] Zhang C, Li J, Wang H, et al. Evaluation of dynamic optical service restoration on a large-scale ROADM mesh network[J]. IEEE Communications Magazine, 2019, 57(4): 138-143.
- [3] X. Liu, N. Deng, Emerging optical communication technologies for 5G, in Optical fiber telecommunications VII, pp. 751-783.
- [4] N. Deng, ROADM/OXC and related technologies/standards: enabling more agile optical network, invited talk at market focus, ECOC 2017.
- [5] Li Y, Zong L, Gao M, et al. Colorless, partially directional, and contentionless architecture for high-degree ROADMs, Optical Fiber Communication Conference. Optical Society of America, 2020: M4D. 2.
- [6] Marom, Dan M., et al. Survey of photonic switching architectures and technologies in support of spatially and spectrally flexible optical networking. Journal of Optical Communications and Networking 9.1 (2017): 1-26.
- [7] Li Y, Gao L, Shen G, et al. Impact of ROADM colorless, directionless, and contentionless (CDC) features on optical network performance[J]. Journal of Optical Communications and Networking, 2012, 4(11): B58-B67.
- [8] Way W I. Next generation ROADM architectures[C]//Asia Communications and Photonics Conference. Optical Society of America, 2012: AS1G. 3.
- [9] Zong L, Zhao H, Feng Z, et al. Demonstration of ultra-compact contentionless-ROADM based on flexible wavelength router[C]//2014 The European Conference on Optical Communication (ECOC). IEEE, 2014: 1-3.
- [10] Frisken S J. Polarization diverse wavelength selective switch: U.S. Patent 9,654,848[P]. 2017-5-16.
- [11] Zami T, Lavigne B, Balmefrezol E. Crosstalk analysis applied to wavelength selective switches[C]//Optical Fiber Communication Conference. Optical Society of America, 2006: OFP4.
- [12] Yang H, Wang R, Wright P, et al. Impact of WSS passband narrowing effect on the capacity of the flexible-spectrum networks[C]//Optical Fiber Communication Conference. Optical Society of America, 2017: W11. 5.
- [13] Tibuleac S, Filer M. Transmission impairments in DWDM networks with reconfigurable optical add-drop multiplexers[J]. Journal of Lightwave Technology, 2010, 28(4): 557-568.
- [14] Cibby Pulikkaseril, Luke A. Stewart, Michaël A. F. Roelens, Glenn W. Baxter, Simon Poole, and Steve Frisken, Spectral modeling of channel band shapes in wavelength selective switches, Opt. Express 19, 8458-8470 (2011)
- [15] Fabrega, J. M., et al. On the filter narrowing issues in elastic optical networks. Journal of Optical Communications and Networking 8.7 (2016): A23-A33.
- [16] Nelson, L. E., et al. 100Gb/s dual-carrier DP-QPSK performance after WDM transmission including 50GHz Wavelength Selective Switches. 2011 Optical Fiber Communication Conference and Exposition and the National Fiber Optic Engineers Conference 0.
- [17] Scholl T J, Rehse S J, Holt R A, et al. Broadband precision wavelength meter based on a stepping Fabry-Pérot interferometer[J]. Review of Scientific Instruments, 2004, 75(10):3318-3326.
- [18] A. Matsushita, M. Nakamura, K. Horikoshi, S. Okamoto, F. Hamaoka and Y. Kisaka, 64-GBd PDM-256QAM and 92-GBd PDM-64QAM Signal Generation using Precise-Digital-Calibration aided by Optical-Equalization, 2019 Optical Fiber Communications Conference and Exhibition (OFC), 2019, pp. 1-3.
- [19] H. Li et al., Real-time Demonstration of 12-λ×800-Gb/s Single-carrier 90.5-GBd DP-64QAM-PCS Coherent Transmission over 1122-km Ultra-low-loss G.654.E Fiber, 2021 European Conference on Optical Communication (ECOC), 2021, pp. 1-4.
- [20] K. Seno, K. Suzuki, K. Watanabe, N. Ooba and S. Mino, Channel-by-channel tunable optical dispersion compensator consisting of arrayed-waveguide grating and liquid crystal on silicon, OFC/NFOEC 2008 - 2008 Conference on Optical Fiber Communication/National Fiber Optic Engineers Conference, 2008, pp. 1-3.



New Architecture of x-Dimension Optical Access

Kuo Zhang, Zhicheng Ye, Borui Li, Liangchuan Li
Optical Research Dept

Abstract

In the F5G era, optical access networks start to extend from home broadband to various industry scenarios. New industrial applications are more demanding of the quality and reliability of optical access networks and put great strain on the traditional time division multiple access (TDMA) mechanism which has been successful in fixed access networks. x-Dimension Optical Access (xDOA) aims to explore a new underlying optical access architecture for the future, and develop new multiple access multiplexing and multi-dimension resource scheduling technologies to meet the requirements of premium networks and lead the optical access field along a new path. First, as a new multiple access technology, xDOA introduces the optical domain multiple access coherent digital signal processing algorithm, which adopts the advantages of time division multiple access (TDMA) to fully share the bandwidth and makes full use of the high-quality feature of frequency division multiple access (FDMA) hard pipes to eradicate problems of the existing passive optical networks (PONs). Second, xDOA greatly simplifies the optoelectronic front end of optical network units (ONUs) through an innovative multiple access coherent architecture. Specifically, it matches the service rate with the component bandwidth through frequency selection on receivers, thereby reducing the requirements for the ONU component bandwidth and reducing costs for users. Third, xDOA not only reshapes the multiple access multiplexing technology on the physical layer, but also introduces the multi-dimension scheduling technology on the media access control (MAC) layer. The two technologies are used together to make service bearing more flexible and robust through more diversified multi-parameter monitoring. Fourth, in xDOA, the new MAC layer identifies quality of service (QoS) requirements of different service types and schedules resources to pipes with different priorities. Flexible scheduling technologies are used to isolate services at different levels, meeting various service requirements in the P2MP architecture and improving the quality of the entire network.

Keywords

new optical access architecture, optical domain multiple access technology, multi-dimension scheduling technology, the fifth generation fixed network (F5G)

1 Introduction

With the continuous societal and economic development, people have ever increasing requirements for the bandwidth and quality of communications networks. Optical fiber access, with its high bandwidth, has become the most in-demand technology and future-proof solution for the last-mile access network. In addition, China's 14th Five-Year Plan and 2035 Vision Outline clearly states the goal of "promoting gigabit optical networks". This marks the first time for gigabit optical network to be written into the government's work report. Furthermore, the European Telecommunications Standards Institute (ETSI) officially established the fifth-generation fixed network (F5G) organization, which proposed "Gigabit optical networks" [1]. Fiber access is becoming the new foundation of the digital economy and will boost the development of multiple industries, such as IT, cloud, and manufacturing.

In earlier generations of fixed networks, fiber access is mainly used in home scenarios, and bandwidth is the foremost requirement. The traditional time division multiple access (TDMA) PON architecture can work with dynamic bandwidth allocation (DBA) to meet home bandwidth requirements. The TDMA-based optical access network was extremely effective over the last two decades due to its bandwidth sharing and statistical multiplexing features [2, 3]. However, in the F5G era, optical access networks have extended from home broadband to various industries, such as campus, manufacturing, healthcare, and finance. Applications in new industries are more demanding of the quality and reliability of optical access networks. As shown in Figure 1, ETSI defines three features of F5G: enhanced fixed broadband (eFBB), supporting high-speed, large-capacity communication, and high spectral efficiency; full-fiber connection (FFC), meeting service requirements for extensive coverage, connectivity extension, and massive-data dense communication; and guaranteed reliable experience (GRE), ensuring low-latency, high-reliability, high-availability, and high operating efficiency [1]. To meet these new requirements, traditional PONs are facing multiple challenges. In TDM PON user access control, multiple users share the same channel at the physical layer. This feature maximizes bandwidth utilization, but has shortfalls when it comes to latency, burst, rogue ONT, hard pipe, and reliability. Such shortfalls prevent PON from meeting various industrial requirements. For example, in the industrial field, production and non-production services need to be strictly isolated, and latency-sensitive services require

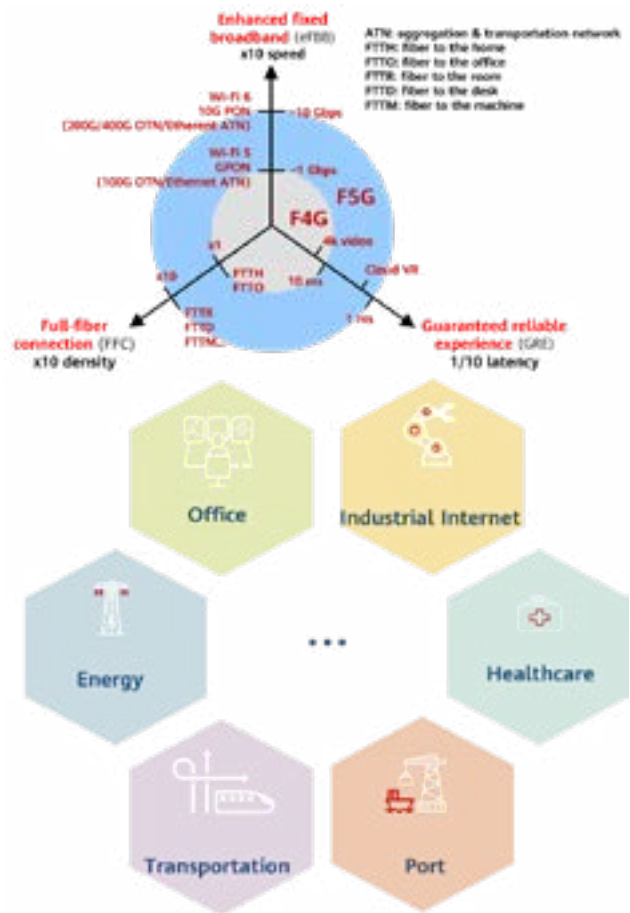


Figure 1 F5G features and new optical access scenarios

deterministic low latency. Meanwhile, network security is a basic requirement in the financial and government sectors. The service level agreement (SLA) requirements of these services pose serious challenges to current optical access networks. To address challenges faced by the traditional PON, we optimized the protocol and software to find differentiated characteristics and improve the quality of the PON. For example, network slicing was used to isolate services, software was used to detect rogue devices, and single-frame multi-burst function was used to reduce the latency [4, 5]. These technologies meet the requirements of premium networks and certain new service scenarios to some extent. However, the TDM-PON is still based on TDMA, and cannot fully meet the requirements of bearing different services.

In addition, the rate increase in each generation of PON result in the decrease of power budget and sensitivity. Theoretically, the sensitivity decreases by at least 3 dB each time the rate doubles. Currently, in all reported 50G TDM-PON sensitivity tests, -27 dBm is already the limit [6-8]. This means that the transmitter needs to have at least 8 dBm output optical power to maintain a 35 dB power budget.

According to previous experience, the rate increases by four times for each new generation in the evolution process from GPON to XGPON and to 50GPON. We can therefore infer that the next generation following 50G PON will deliver a capacity of 200G, rendering the current architecture obsolete.

To address the preceding challenges in the evolution of traditional TDM-PON, this paper introduces the new xDOA architecture. By exploring future optical access technologies, this paper describes a new underlying architecture and resource scheduling solution to reshape the multiple access and multi-dimension scheduling technologies and meet the requirements of future premium networks, thereby enabling optical access to enter a new era.

2 Multiple Access and Scheduling Technologies Related to Optical Access Networks

2.1 Overview of Communication Multiple Access Technologies

In a point to multipoint (P2MP) communication scenario, the multiple access multiplexing technology is an important research direction and discipline. Typical technologies include time division multiple access (TDMA), frequency division multiple access (FDMA), code division multiple access (CDMA), non-orthogonal multiple access (NOMA), and other such technologies. Multiple access technologies are widely used at the access side of a network. For example, in the field of radio and Wi-Fi communications, each generation of evolution means the application of a new multiple access multiplexing technology, which can be TDMA, CDMA, FDMA, or orthogonal FDMA (OFDMA).

PON is a multi-point communications network widely used in the optical communications field. As shown in Figure 2, the past 20 years have seen PON systems evolve from GPON to 10G PON and then to 50G PON. During this time, PON has been based on TDMA. A new multiple access technology (xDMA) is now in the academic research stage. For one thing, optical access scenarios were limited and requirements were relatively simple in the past. For another, a new multiple access multiplexing technology depends on the maturity of digital chips with a high-bandwidth and sampling rate. Digital chips for optical communications gradually matured with development of coherent optical communications in

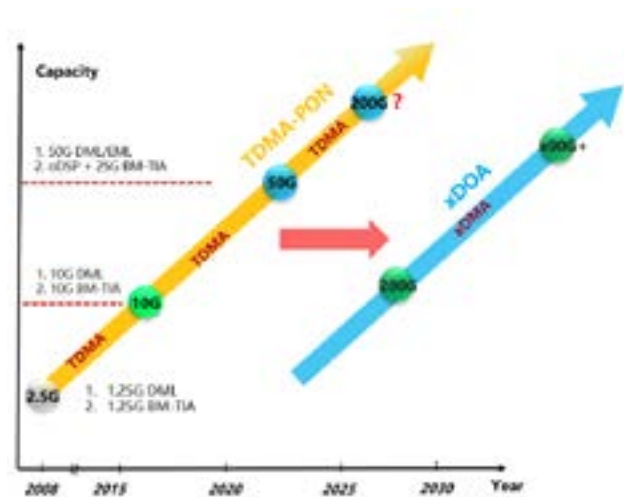


Figure 2 New xDOA architecture

the last ten years. Therefore, the research on the new xDOA architecture is to promote the transformation of core optical access technologies and thereby meet the requirements of new scenarios and industries and define the future optical access architecture.

2.2 Advantages of and Challenges Faced by TDM-PON

Since the emergence of PON technology in the 1990s, the TDMA technology has formed the foundation of the PON architecture. Figure 3 shows the typical architecture of TDM-PON. At the optical network unit (ONU) transmit end, different ONUs control the switches of directly modulated lasers (DMLs) by using a burst mode driver (BM-driver) to implement multiplexing of different timeslots. After passing through the optical distribution network (ODN), multiple time division optical signal channels are combined into one channel to send signals to the receiver of the optical line terminal (OLT). At the OLT receive end, different burst signals have different amplitudes due to differences between ONU transmission distances and link insertion losses, and protection time needs to be reserved between burst signals. This requires analog components, such as the trans-impedance amplifier (TIA) and clock and data recovery (CDR) of the receiver, to have the burst receiving function. The TIA needs to provide fast gain response to pull up burst signals with a low amplitude and equalize amplitudes of different ONU signals. The CDR needs to be able to quickly lock a clock and a phase to shorten the convergence time and reduce burst overheads.



Figure 3 Typical TDM-PON architecture

Judging from past experience, TDM-PON has several outstanding advantages that have led to its success in broadband access. First, a TDM-PON has a relatively low requirement on the frequency stability of a laser, and different burst frames are completely isolated in terms of time, thereby effectively avoiding crosstalk. For this reason, a laser in the TDM-PON is usually an uncooled distribution feedback (DFB) laser with direct modulation. As the laser does not require thermo electric cooler (TEC) control, the price is very low. In addition, the TDMA architecture can implement sharing and statistical multiplexing of bandwidth resources between different ONUs, thereby maximizing bandwidth utilization. Third, in the early stage, transceiving of the TDMA architecture can be implemented with just an analog optoelectronic device, without requiring a digital chip. This has good feasibility when high-speed digital optical communications chips are not mature.

However, TDMA also faces many challenges. First, the traditional PON network implements multi-user access based on TDMA, the upstream communication depends on burst mode optoelectronic components, and there are technical bottlenecks in the evolution to higher speeds. Second, in a TDM-PON, multiple users share the same channel at the physical layer by means of time division. Although this feature can maximize bandwidth sharing, bandwidth resources cannot be exclusively used at the physical layer, making it difficult to implement hard pipes at the physical layer. Third, in a TDM-PON, the optical physical layers of ONUs and OLTs need to have the same data processing capability. For example, in a 10G PON network, an ONU needs to have the 10 Gbit/s transmit and receive capability, but the actual service rate of an ONU may be only 1 Gbit/s, which means that bandwidth is wasted by ONU components. Fourth, the evolution of each generation of PON technologies requires re-planning of wavelengths and backward compatibility through multiple combo transceivers. For example, the XG PON combo optical module needs to be compatible with the GPON transceiver in order to support the coexistence of XG PON and GPON, and 50G PON optical

modules needs to be compatible with XG PON sending and receiving. In the future, 100G/200G PON may need to be compatible with 50G PON and XG PON. This is not only technically difficult, but also uneconomical.

2.3 Network Scheduling Technologies and Challenges Faced by Optical Access Networks

In an optical access network, bandwidth requirements vary both between users and for the same user at different times. To meet such diversified bandwidth requirements, an OLT allocates bandwidth according to the service conditions of each ONU and the size of buffered data. In this way, the bandwidth resources of the network are fully utilized. Common bandwidth allocation modes include dynamic bandwidth allocation (DBA) and fixed bandwidth allocation (FBA). With bandwidth increasingly becoming a major selling point for operators over the past two decades, DBA can be used to classify bandwidths and provide bandwidth on a per-user basis, fully meeting user requirements.

With the development of optical access networks in various industries, simply increasing bandwidth cannot fully meet user requirements. For example, in the industrial field, latency and jitter are important factors for customers. In the enterprise field, customers are more concerned about reliability and security. Although the P2MP architecture is inherently capable of bearing multiple services, traditional DBA cannot meet QoS requirements of different service types. For example, when to home (2H) and to business (2B) PON ports are used together, DBA cannot distinguish the service types. As a result, home broadband services may preempt the bandwidth of enterprise services. For some industry services, production and non-production services need to be isolated. The P2P system architecture is costly, and the P2MP architecture cannot meet the isolation requirements of different services due to the limitation of the scheduling solution. Therefore, in the future, optical access networks will be classified not only by bandwidth, but also by latency, security, reliability, and other specifications. The scheduling technology of the optical access network will gradually evolve from DBA to DxA, thereby implementing multi-dimension scheduling and meeting the requirements of F5G and even F6G.

3 Directions for New xDOA Architecture Design

As shown in Figure 4, to cope with existing challenges, xDOA needs to reshape the current TDM-PON architecture and resolve technical bottlenecks of underlying TDMA from the following aspects:

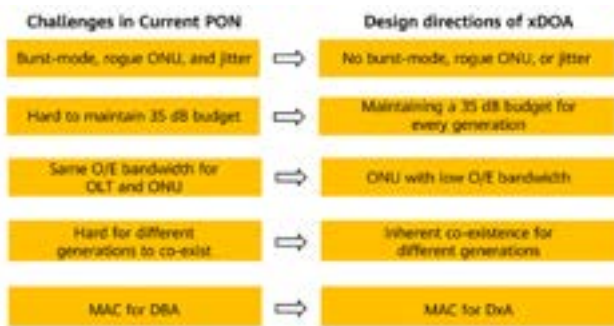


Figure 4 Design direction of the xDOA architecture

First, the new architecture needs to support concurrent access by multiple ONUs to avoid the burst problem in the current system and in doing so break the bandwidth bottleneck faced by a burst component in the evolution to higher speeds. Second, in the future, optical access should be able to eradicate rogue ONTs and jitter, and provide premium services for high-value users. Third, the optoelectronic bandwidth at an ONU needs to match the service rate. That is, a GHz component can be used to implement service access at a gigabit rate, so as to reduce the ONU cost. This is particularly important in a high-speed PON system. For example, when the PON system capacity exceeds 50G, if the effective service rate of each ONU is only several gigabits, using 50 GHz components at the ONU will be a waste of costs.

The new xDOA architecture must also solve the long-term evolution problems of optical access in the future. First, the ODN construction cost is high and re-work must be avoided, which means that power budget of a PON system must have a sufficient margin. Given this, the new xDOA architecture must have high theoretical sensitivity and power budget to support long-term architecture evolution. Second, new wavelengths must be planned for each generation of PON to achieve backward compatibility. The PON system has high upgrade costs and slow inter-generation evolution, generally taking 10 years. The new xDOA architecture must support smooth evolution and reduce the cost of multi-generation and multi-rate co-existence, in addition to supporting rapid incremental evolution. Finally, the MAC layer of the new

xDOA architecture must be able to flexibly schedule resources based on the conditions and requirements of different services. In addition to allocating different bandwidths (timeslots) based on traditional traffic monitoring and buffer reports, the new MAC layer need to identify service types and bear services on different pipes based on the QoS requirements. In addition, the MAC layer needs to coordinate with the physical layer to properly allocate resources based on the changes of service bearer pipe conditions, make full use of system channel features, improve system robustness, and enhance system maintainability, manageability, and intelligent maintenance capabilities.

4 New xDOA Architecture

4.1 Overview of the New xDOA Architecture

Figure 5 is an overall block diagram of the new xDOA architecture. In general, the architecture is based on FDMA and simplified coherent transceiving technologies. In the downstream direction, an OLT transmits multiple subcarrier signals. An ONU adjusts the wavelength of the local oscillator laser, and receives a specific subcarrier by means of coherent frequency selection, thereby implementing FDMA. In the upstream direction, ONUs use different subcarriers, and the OLT receives subcarriers from the ONUs by means of coherent detection. In the new xDOA architecture, different ONUs can use independent subcarrier channels in the upstream direction to implement continuous transmission and reception. Unlike in TDMA mode, ONUs do not need to wait for their own timeslots. In this way, hard pipes are achieved to meet bandwidth and latency requirements. In the downstream direction, ONUs only need to receive their own data at the physical layer, reducing the optoelectronic bandwidth requirements on ONUs.

The FDMA architecture of xDOA also supports the TDMA technology. In TDMA mode, each subcarrier is shared by multiple ONUs, realizing time-frequency hybrid multiple access multiplexing. As shown in Figure 6, the bandwidth resource is divided into different time-frequency resource blocks. Each ONU completes data transmission using a separate resource block. The time-frequency hybrid multiple access mode provides many benefits. First, ONUs can be classified into different types, and an independent subcarrier can be allocated to an ONU that has a high reliability and latency requirement, for example, ONU

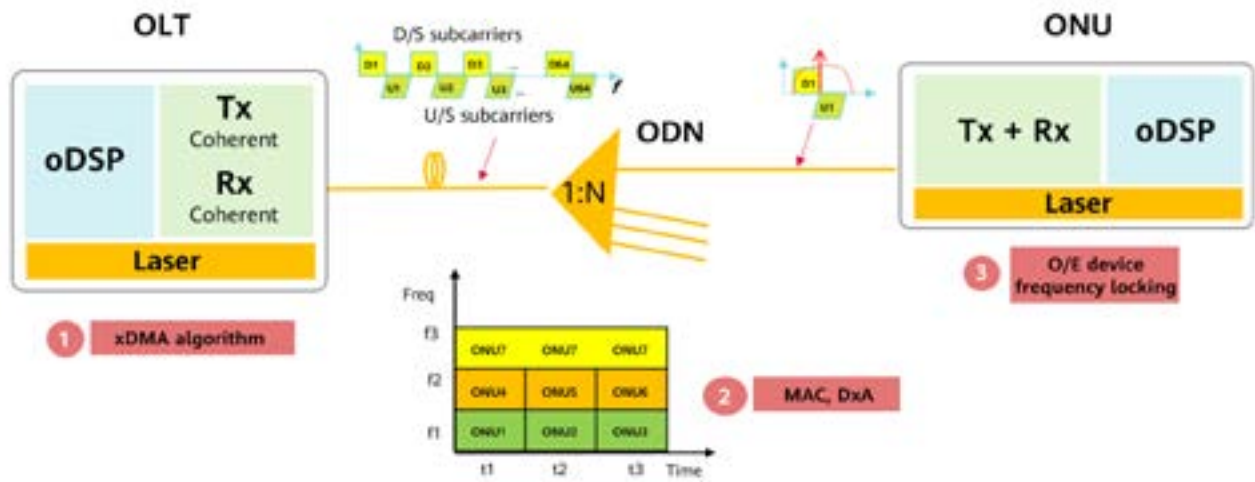


Figure 5 Block diagram of the xDOA architecture

10 in Figure 6. Multiple regular ONUs can share one subcarrier, for example, ONUs 1 to 9 in the figure. In other words, in xDOA, resource allocation is based on both bandwidth requirements and specifications such as latency and reliability. Second, time-frequency hybrid multiple access can make full use of the flexible multiplexing advantages of TDMA as well as the hard pipe advantages of FDMA, thereby meeting the requirements of different applications and scenarios and maximizing the utilization of network resources. Third, TDMA reduces the number of subcarriers and increases the subcarrier spacing under the same capacity, thereby lowering the frequency stability requirement on the ONU-end transmitter laser and local oscillator laser.

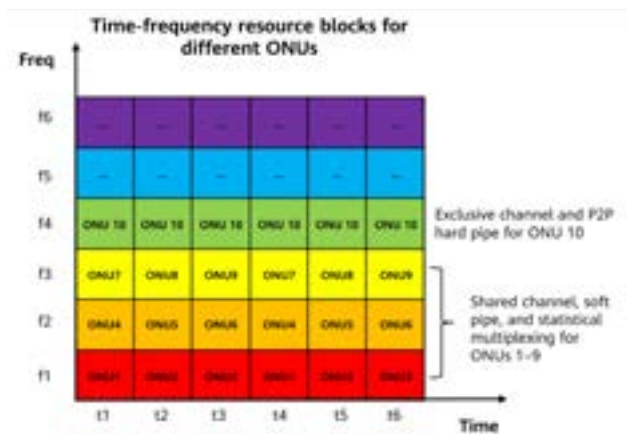


Figure 6 FDMA+TDMA time-frequency resource blocks in the optical domain

4.2 Advantages of the New xDOA Architecture

The new xDOA architecture has the following advantages:

First, the xDOA multiple access coherent technology effectively improves receiver sensitivity. In coherent optical communication, the signal light is amplified by the local oscillator light of the coherent receiver, which can effectively improve the receiver sensitivity. For example, for 10 Gbit/s signals, the sensitivity of a coherent receiver can be higher than -45 dBm, but the sensitivity of an APD-based receiver is only about -30 dBm. This inherent advantage in sensitivity ensures that the optical power budget is kept above 35 dB while the rate is increased.

Second, xDOA can fundamentally solve the burst and jitter problems. When FDMA is used, subcarriers are completely isolated, and an ONU can exclusively use a hard pipe without jitter. In addition, simultaneous transmission of multiple ONUs can ensure relatively stable optical signals at the OLT receive end, reducing the requirement on the burst mode component.

Third, xDOA reduces dependence on optoelectronic devices and prevents ONU devices from wasting bandwidth. Coherent communication has an inherent electrical frequency selection function. An ONU only needs to receive a single subcarrier, and the corresponding physical-layer optoelectronic detector, trans-impedance amplifier, and oDSP requires only the bandwidth capability of a single subcarrier. For example, for an 8-carrier 64 GHz system,

each ONU only needs to receive the 8 GHz subcarrier. This principle is illustrated in Figure 7. In a traditional PON network, the optical physical layer of an ONU needs to receive all data, after which the MAC module of the ONU selects its own data. In xDOA, the optical physical layer of an ONU receives only a specific subcarrier, and undertakes some functions of the MAC module. This greatly reduces power consumption and chip surface area.

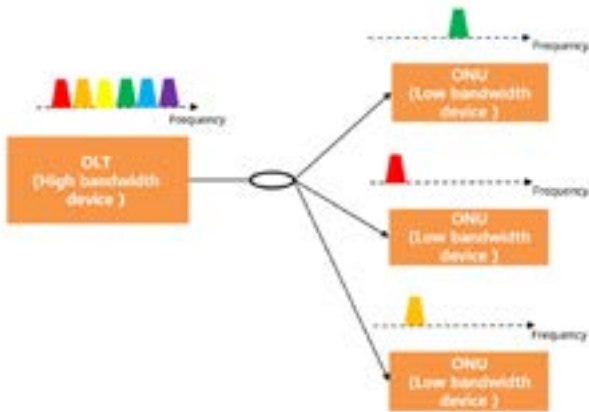


Figure 7 OLTs and ONUs with asymmetrical optoelectronic devices in xDOA

Fourth, xDOA can reduce the cost of rate evolution between different generations by using flexible modulation and reception technologies. The digital coherent technology enhances the architecture flexibility and inherently supports the co-existence of multiple generations and rates. As shown in Figure 8, all information of optical signals can be flexibly modulated. For example, different modulation formats and constellation shaping technologies can be used to achieve

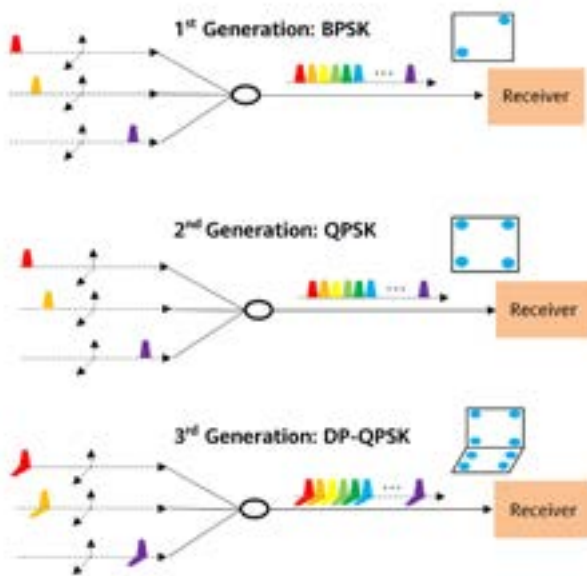


Figure 8 Rate upgrade and evolution of the xDOA architecture

flexible rates. Another way flexible rates is achieved is by flexibly selecting dimensions such as polarization, phase, and amplitude. Optical signals of different rates can coexist at the same wavelength, thereby effectively avoiding technical bottlenecks faced by the multi-wavelength combo technology in a multi-generation PON system.

Fifth, xDOA time-frequency resource division provides a foundation of a physical pipe for DxA multi-dimension scheduling, and provides hard pipes or shared pipes based on user bandwidth, latency, and security requirements.

5 Key Technologies of the xDOA Architecture

5.1 Multiple Access Coherent Architecture and Algorithm

As an important optical communications domain, multiple access multiplexing is a nascent technology. In the previous part of this paper, we introduced the idea of combining FDMA and TDMA. In fact, there are many other new multiple access technologies (xDMA) in the optical domain, such as CDMA, OFDMA, and NOMA. Applying these technologies in multiple access coherent optical communications is also an important topic.

Compared with direct detection, coherent optical communication has a relatively complex transceiving structure. Each transmitter includes four Mach-Zehnder modulators (MZMs), and each receiver includes four bit-plane decoders (BPDs). In addition, there are optical mixers, polarization beam combiners and splitters, and the like. An access network is oriented to end users, and is sensitive to costs. Therefore, simplifying the coherent transceiving architecture is also a key topic. For example, we can use innovative detection methods to reduce the number of optoelectronic components on the ONU side, and use optical integration methods to reduce the size of optical modules or components. In addition, while simplifying the coherent transceiving architecture, the current optical communications industry chain needs to be reused as much as possible to minimize costs. In recent years, universities and research institutes around the world have conducted some research in this field [9–17]. For example, to simplify the coherent transceiving architecture, it has been found that 3 x 3 couplers can be used to simplify the receiver complexity, the number of receivers can be reduced by code

heterodyne detection at the transmit end, and the analog coherent technology can be used. These are all subjects worth further study.

The multiple access multiplexing algorithms are the key technologies for a multiple access coherent architecture. For example, in a time-frequency hybrid multiple access architecture, the subcarrier spacing compression algorithm, inter-subcarrier crosstalk cancellation algorithm, and time division signal fast receiving algorithm are all critical. In addition, although multiple access algorithms are widely applied in fields such as Wi-Fi and radio communications, the possibility of combining them with optoelectronic devices in the optical communications field is also a subject worth studying. For example, in radio communications, a radio frequency antenna generally works in a GHz frequency band, but optical communications work in the 190 THz frequency band. The radio frequency of local oscillator signals is also relatively stable, and the phase noise is relatively small. However, the frequency offset of a communication laser can reach the GHz level, and the phase noise can reach the MHz level. All such factors affect research and selection of multiple access algorithms. On the other hand, for a time-frequency hybrid multiple access system, quickly locking upstream time division signals is also a key technology. This involves technologies such as fast clock recovery, fast SOP tracking, and fast frequency offset and phase noise estimation.

5.2 Frequency-Locking Laser and Simplified Optoelectronic Components

In the FDMA optical access architecture, optical signals are transmitted by multiple subcarriers divided by spectrum. Subcarriers occupy different frequencies and are allocated to different ONUs. This architecture requires a high-precision optical frequency locking mechanism to avoid signal crosstalk between adjacent subcarriers. For example, assuming that each subcarrier occupies 8 GHz bandwidth and the subcarrier spacing is 9 GHz, avoiding signal crosstalk between subcarriers requires the frequency precision of the laser to reach ± 500 MHz. In traditional coherent optical communication, an external cavity laser (ECL) is generally used together with Etalon wavelength locking to improve the optical frequency precision and stability of a laser.

However, a laser is costly and takes up the majority of the optical module cost, which is unacceptable in cost-sensitive optical access networks. For this reason, in xDOA, we need to find out a laser solution that features high-precision and low-cost frequency locking.

In a wavelength division multiplexing (WDM) system, the wavelength of a laser needs to cover the entire C band, the tuning range is large, and costs are difficult to drive down. An optical access system has only two wavelengths: an upstream wavelength and a downstream wavelength. Also, the tuning range of a laser only needs to cover the signal spectrum width, which is usually about 1 nm. For example, for an 8-subcarrier system, if the width of each subcarrier is 8 GHz, the tuning range of a subcarrier only needs to cover 64 GHz. Therefore, a DFB laser based on temperature tuning can meet the tuning range requirement, which greatly reduces the cost of lasers.

However, the wavelength precision of a common DFB laser is usually several GHz or even dozens of GHz, which cannot meet coherent multiple access requirements. Therefore, we need to consider the laser frequency locking solution from the system and technology perspectives to meet requirements of the new architecture.

In a PON system, customers are highly conscious of ONU costs. Simplifying the optoelectronic front ends of ONUs in a multiple access coherent system is one of key technologies. For one thing, optoelectronic components of ONUs need to be simplified in the multi-access coherent architecture design to implement simplified transceiving on ONUs. For another, the cost needs to be reduced at the component level. Important research directions include silicon photonics integration, joint optoelectronic design, and low-drive-voltage optical modulator design. For example, in terms of silicon photonics integration, we need to pay special attention to the insertion loss and drive voltage of silicon photonics modulator to ensure that the output optical power of ONUs meets the requirement of the optical power budget. Further to this, we can simplify optoelectronic components from the system level. For example, due to the coherent frequency selection feature of the xDOA multi-subcarrier system, the xDOA has a low requirement on the bandwidth of ONU optoelectronic components, and ONUs do not have a high requirement on the size of optoelectronic components. We can leverage these features to optimize the design of ONU optoelectronic components.

5.3 Collaboration Between DxA and Multiple Access

Traditional optical access networks are mainly oriented to home broadband. Bandwidth access and efficient statistical multiplexing of multiple users are the main factors to be considered. One of the key MAC technologies of traditional optical access networks is dynamic bandwidth allocation (DBA). The xDOA architecture, with a larger power budget, more connections, and more diversified pipes, will carry different services in various scenarios, such as production and non-production services in industrial and campus networks. DxA is a technology that provides reasonable, efficient, and robust resource allocation for different services.

In a traditional TDM PON, the DBA technology is decoupled from TDMA at the physical layer. The DBA allocates resources only by detecting the traffic of different users and reporting the buffer status of different traffic. Therefore, its pipes are rigid and not robust. Once the channel condition changes, users will be forced to go offline. In an xDOA system, the physical layer has both time and frequency resource blocks. Therefore, the DxA and the xDMA need to be correlated. In addition to monitoring of the foregoing two parameters, an oPHY signal processing chip of the xDMA provides more auxiliary information for DxA, such as SNR changes of different subcarriers and pre-FEC bit error redundancy. DxA flexibly allocates resources based on the information. For example, more flexible coding and modulation are performed on user information in different timeslots of a subcarrier to match a channel condition, or switch services to another channel due to deterioration

of channel conditions in a frequency domain. This flexible cross-layer orchestration enables user services to better match ODN conditions, and also makes user access more robust. When ODN conditions deteriorate or an optical transceiver deteriorates, multi-parameter monitoring and flexible resource allocation are performed to ensure high user access quality.

5.4 Collaboration Between DxA and TM

Due to the wide deployment of ODNs and the ubiquitous P2MP connections, hybrid access of different services is inevitable in an optical access system. The xDOA system is bound to carry services for various industries. Therefore, the DxA technology and traffic management (TM) need to work together to properly allocate resources to services to meet different service level agreement (SLA) and quality of service (QoS) requirements.

In some scenarios, 2B and 2H services, or production and non-production services are accessed concurrently. However, 2B services have higher requirements on security and isolation than 2H services, and production services have higher requirements on latency and jitter than non-production services. For example, in industrial and campus scenarios, an independent subcarrier can be allocated to users who have high requirements on latency and reliability. As shown in Figure 6, ONU 10 exclusively occupies a subcarrier, and its bandwidth and latency are not affected by other ONUs. In this way, hard pipe experience is achieved. Therefore, the xDOA system needs to identify the service type and priority. After the number of users connected to a port increases gradually, TM and DxA need to collaborate to flexibly adjust the pipe of each user based on the user and service priority, to ensure isolation, high security, and premium user experience in high-level services. In xDOA, we need to design scheduling and TM mechanisms based on bandwidth, latency, and reliability to provide differentiated network experiences for users and meet the requirements of all-optical connections in various scenarios.

In the future, with the development of fiber to the room (FTTR) technologies and deep integration of optical access networks and Wi-Fi networks, PONs and Wi-Fi networks can be jointly scheduled and optimized to improve the latency and throughput of the entire network. This will also become an important research direction of multi-dimension scheduling.

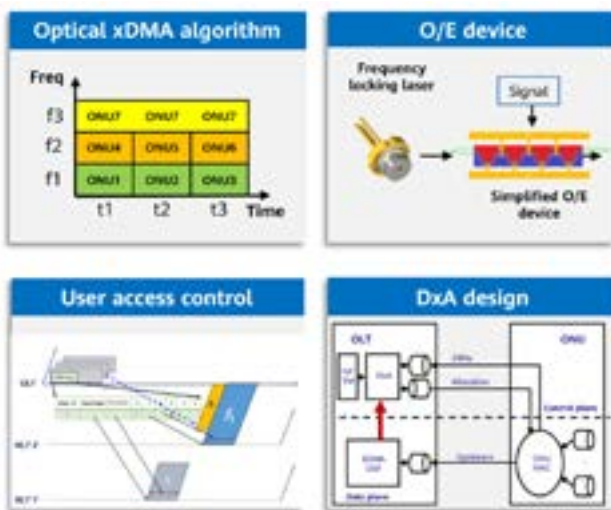


Figure 9 Key xDOA technologies

6 Summary and Prospects

This paper describes the new xDOA architecture and key technologies of this architecture. With innovations in multiple access technologies at the physical layer and multi-dimension scheduling at the MAC layer, this architecture attempts to resolve challenges brought by new applications to optical access networks, implement hard pipes physically, and achieve deterministic low latency to meet the requirements of optical access in various industries. Looking back, optical access networks with TDMA and DBA as core technologies were very successful in meeting the bandwidth requirements of users. In the future, the xDOA architecture with xDMA and DxA as core technologies will reshape the next-generation optical access and become the main technical path of future optical access.

References

- [1] "The Fifth Generation Fixed Network (F5G): Bringing Fibre to Everywhere and Everything", ETSI White Paper No. #41.
- [2] "10-gigabit-capable passive optical networks (XG-PON), " ITU-T Recommendation G.987 series.
- [3] D. Nessel, "PON roadmap, " J. Opt. Commun. Netw., vol. 9, no. 1, pp. A 71–A76, 2017.
- [4] S. Zhou, X. Liu, F. Effenberger, and J. Chao, "Mobile-PON: A High-Efficiency Low-Latency Mobile Fronthaul Based on Functional Split and TDM-PON with a Unified Scheduler," in Optical Fiber Communication Conference, OSA Technical Digest (online) (Optica Publishing Group, 2017), paper Th3A.3.
- [5] S. Hatta, N. Tanaka, and T. Sakamoto, "Low latency dynamic bandwidth allocation method with high bandwidth efficiency for TDM-PON, " NTT Tech. Rev., vol. 15, no. 4, pp. 1–7, 2017.
- [6] L. Zhou and D. Liu, "Symmetric 50G PON using NRZ, " March 2018 IEEE 802.3 Plenary, Chicago, IL, USA
- [7] D. Nessel, et al "50G PON ONU Rx Sensitivity," contribution 21-02-08 Zoom, D10, ITU-T SG15/Q2 (2021).
- [8] M. Tao, J. Zheng, X. Dong, K. Zhang, L. Zhou, H. Zeng, Y. Luo, S. Li, and X. Liu, "Improved dispersion tolerance for 50G-PON downstream transmission via receiver-side equalization, " in Optical Fiber Communication Conference (OFC) (Optical Society of America, 2019), paper M2B.3.
- [9] D. Lavery, S. Erkilinc, P. Bayvel, and R. I. Killey, "Recent progress and outlook for coherent PON, " in Optical Fiber Communication Conference (2018), paper M3B.1.
- [10] D. van Veen and V. Houtsma, "Strategies for economical next-generation 50G and 100G passive optical networks [Invited]," J. Opt. Commun. Netw. 12, A95-A103 (2020)

- [11] J. Zhang, J. S. Wey, J. Shi, and J. Yu, "Single-wavelength 100-Gb/s PAM-4 TDM-PON achieving over 32-dB power budget using simplified and phase insensitive coherent detection, " in European Conference on Optical Communication (2018).
- [12] Dave Welch, Antonio Napoli, Johan Bäck, Warren Sande, João Pedro, Fady Masoud, Chris Fludger, Thomas Duthel, Han Sun, Steven J. Hand, Ting-Kuang Chiang, Aaron Chase, Atul Mathur, Tobias A. Eriksson, Mats Plantare, Magnus Olson, Stefan Voll, and Kuang-Tsan Wu, "Point-to-Multipoint Optical Networks Using Coherent Digital Subcarriers," *J. Lightwave Technol.* 39, 5232-5247 (2021)
- [13] Yixiao Zhu, Lilin Yi, Bo Yang, Xingang Huang, Jun Shan Wey, Zhuang Ma, and Weisheng Hu, "Comparative study of cost-effective coherent and direct detection schemes for 100 Gb/s/ λ PON," *J. Opt. Commun. Netw.* 12, D36-D47 (2020)
- [14] M. S. Erkilinc, D. Lavery, K. Shi, B. C. Thomsen, P. Bayvel, R. I. Killey, and S. J. Savory, "Polarization-insensitive single-balanced photodiode coherent receiver for long-reach WDM-PONs, " *J. Lightwave Technol.* 34, 2034-2041 (2016).
- [15] A. Shahpari, R. M. Ferreira, R. S. Luis, Z. Vujicic, F. P. Guiomar, J. D. Reis, and A. L. Teixeira, "Coherent access: a review, " *J. Lightwave Technol.* 35, 1050-1058 (2017).
- [16] B. Glance, "Polarization independent coherent optical receiver," *J. Lightwave Technol.* 5, 274-276 (1987).
- [17] C. Xie, P. J. Winzer, G. Raybon, A. H. Gnauck, B. Zhu, T. Gersler, and B. Edvold, "Colorless coherent receiver using 3 x 3 coupler hybrids and single-ended detection," *Opt. Express* 20, 1164-1171 (2012).



ACTN Framework for All-Optical Network Automation and Intent-based Applications

Henry Yu, Christopher Janz

Abstract

Abstraction and Control of Traffic Engineered Networks (ACTN) is an SDN solution proposed by IETF. The standard contains both the software architecture and the YANG service models. This paper shows how ACTN, as an architectural solution, can be applied to network automation, which includes the automation of service fulfillment and service assurance. Furthermore, the paper shows that with all-optical access and transport network convergence and new technologies, such as bandwidth optimized networking and small granularity switching with optical service unit (OSU) containers, ACTN can be used to further develop Intent-based and fully-automated E2E applications (or use cases), including as all-optical slicing and all-optical cloud private line. Finally, the paper proposes that, by leveraging network automation technology and Intent-based applications, network operators can build a new management and control infrastructure which integrates network, computing, storage, and user data resources and offers social-network-as-a-service. An ACTN-based solution for the proposed infrastructure is discussed in this paper.

Keywords

ACTN, intent, IBN, SDN, network automation

1 Introduction

The technology advancements in fifth-generation (5G), Cloud, mobile-edge computing (MEC) and other areas are enabling the digitalization of our society [1], as well as the transformation into the digital economy [2]. This transformation shows no signs of slowing down. The anticipated sixth-generation (6G), for example, will enable use cases [3] such as holographic telepresence, remote healthcare, industrial automation, and unmanned mobility, which will have an even more profound impact on our lives and our society.

However, these future use cases also demand a new set of requirements for network performance, architecture, and operations. In particular, three requirements are of paramount importance. First, the complexity of future networks makes manual network configurations and operations infeasible. Thus, network automation becomes an essential requirement [3]. In an autonomous network [4], not only must the network services be configured and deployed automatically in accordance with the user requirements (or intents), but, more importantly, the delivered services must also be automatically assured by closed-loop operations to prevent service disruptions. Equally important, the network performance needs to be monitored automatically and preemptive actions should be taken to prevent network abnormality and service deterioration.

Second, future network services not only need to be ultra-reliable, low latency, high throughput, and energy and cost efficient, but also need to meet all these requirements simultaneously. For instance, unlike 5G, where tradeoffs may be made among these requirements for different 5G slices, 6G networks require that all these requirements be met *holistically* [5].

Third, future networks will require the convergence of networking and computing services [6]. In other words, future networks demand a new network architecture that integrates network, computing, storage, and data resources to offer a new variety of services required by the digital society.

These three requirements represent the essential characteristics of the next-generation network. This paper discusses the optical network innovations and new architectural solutions that enable the optical network to meet these requirements. The rest of the paper is organized as follows. Section 2 presents the concepts of the Abstraction

and Control of Traffic Engineered Networks (ACTN) framework (an IETF standard) and Intent-based networking (IBN). We discuss how ACTN, in conjunction with IBN, may provide a solution to network automation, which includes the automation of service fulfillment and service assurance. Section 3 is dedicated to the technological innovations in all-optical networks, as well as the ACTN-based use case solutions, which may serve as enablers of the networking and computing convergence. In section 4, an ACTN-based architectural solution to the convergence of networking and computing services is proposed. We show that, by leveraging network automation technology and Intent-based applications provided in Section 2 and 3, network operators can build an all-optical network infrastructure that integrates network, computing, storage, and user-data resources and offers a new variety of integrated networking and computing services, including social-network-as-a-service. Future work is discussed in Section 5, and Section 6 draws the conclusion.

2 ACTN Framework and IBN Concepts for Network Automation

This section contains three parts. We first provide an overview of the ACTN framework, a software defined networking (SDN) solution proposed by IETF. Next, we introduce the concept of IBN, which extends the capability of SDN and is an enabler of network automation. Lastly, we extend the ACTN framework by incorporating the IBN concepts to offer a new architectural solution for network automation.

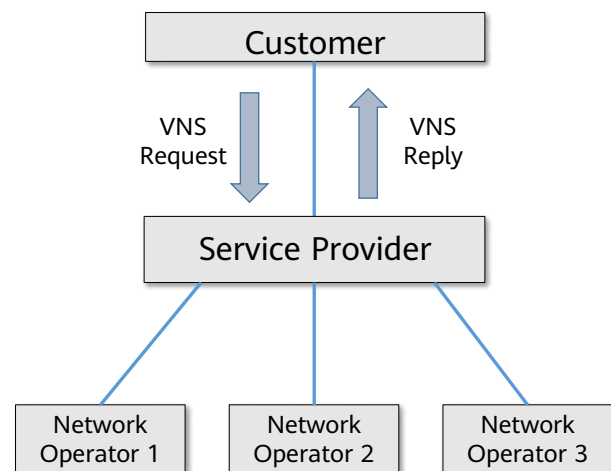


Figure 1 ACTN three-tier business operation model (adopted from [7])

2.1 Overview of the ACTN Framework

The ACTN framework [7] provides a set of management and control functions that operates on TE networks to deliver to customers the virtual networks (VN) that are built from the abstractions of the physical TE networks. Figure 1 depicts the business operation model of ACTN. The customers include businesses, government institutions, residential users. The service provider delivers to the customers Virtual Network Services (VNSs), which are service agreements between a customer and operator to provide VNs. VN [8] is a technology agnostic data model which may serve as a customer's intent. Service providers may or may not be network operators, depending on whether they own the physical networks. Nonetheless, the operation model supports both cases.

The main characteristic of SDN [9] is the combination of a centralized control plane, with a global view of the network resources, and network programmability. ACTN realizes this through a hierarchical controller architecture. The ACTN hierarchical framework consists of three types of controllers, as shown in Figure 2.

A Customer Network Controller (CNC) implements user-level applications that consume the VN services offered by the service provider (or the network operator) via the CNC-MDSC Interface (CMI). The Multi-Domain Service Coordinator translates customer VNS requests to their corresponding network services and provides multi-domain service orchestration and coordination. The MDSC sends the

network service requests to and receives network topology information from the Provisioning Network Controllers (PNCs), via the MSDC-PNC Interface (MPI). A PNC is responsible for configuring the network elements (NEs), monitoring the network topology changes, and collecting topology information.

Because of the architecture's flexibility of being able to deploy multiple heterogeneous PNCs, the ACTN framework is not limited only to management and control of TE networks. It can also control, for example, an IP network by deploying a PNC for the IP domain. In other words, another important characteristic of ACTN is its ability to support a wide variety of layer 0 to layer 3 services. Moreover, IETF provides a rich set of standardized models for these services. They are used in the CMI and MPI, which allows ACTN to offer the customer open and standardized northbound interfaces for network programmability. Figure 3 provides a sample of these service models.

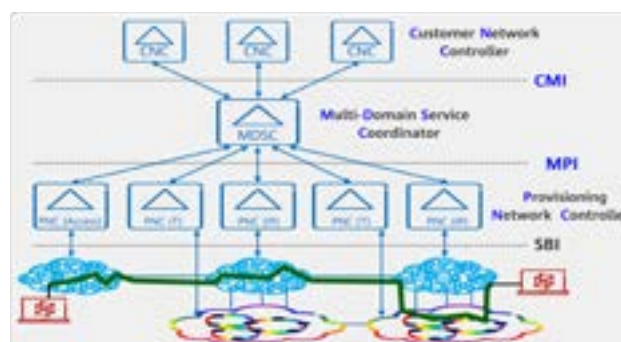


Figure 2 ACTN hierarchical controller architecture (adopted from [10])

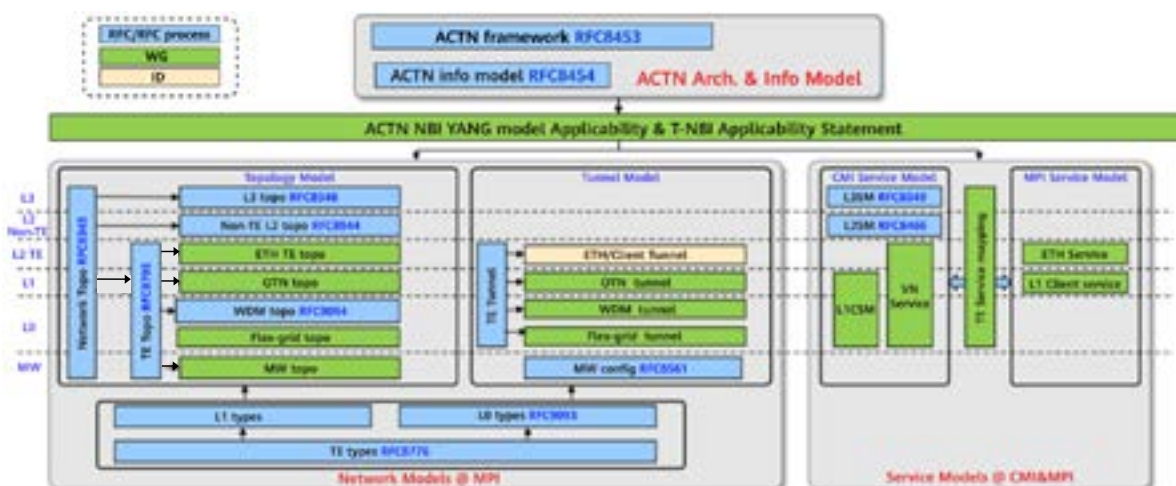


Figure 3 A sample collection of the IETF L0 to L3 service models supported by ACTN (adopted from [10])

2.2 Concepts of IBN

Although SDN is an important step towards the goal of autonomous networks, its limitations are also identified in the process of reaching this goal. Two such limitations are discussed in [11], and how they can be overcome by evolving the SDN solution to the IBN paradigm. The first limitation of SDN is the complexity of programming the network to achieve operational goals using the current SDN northbound APIs, such as TAPI. Furthermore, these APIs are usually technology-specific, and thus require the user to have appropriate domain-specific knowledge. IBN overcomes this limitation by using a different approach. An intent is defined as a declaration of the operational goals that a network should meet and the outcomes that the network should deliver, without specifying how to achieve them [12]. This approach allows the user to define *what* without specifying *how*, and thus reduces the network programming complexity. An IBN system has an Intent Translation functional block, which translates user intents into technology-specific service models. The ACTN framework can be easily extended to support this Intent approach and thus overcome the first limitation. The CMI models can be extended or enhanced to define user intent. In addition, the Intent Translation functional block can be added into MDSC.

The second limitation is SDN's lack of a data analytics-based closed-loop mechanism to support service assurance – a feature required for building an autonomous network. In the IBN paradigm, Intent Assurance is a functionality which ensures that the deployed network services and configurations indeed comply with the desired intent, and the closed-loop mechanism is a means to achieve that. A closed-loop may use technologies such as machine learning (ML) to anticipate network anomalies and service degradations before they take place and thus enable proactive actions for preventing them (e.g., reroute traffic or increase services' maximum bandwidth). ETSI ZSM 009-1 [13] provides a specification for the closed-loop, as depicted in Figure 4.

The closed-loop control flow contains four stages: monitoring, analysis, decision and execution. The green blocks in Figure 4 represent the logical functional entities that carry out the control flow. The ACTN framework may incorporate the closed-loop mechanism as follows.

- M2A (monitoring): The Data Monitoring module collects telemetry and performance data from the network elements and forwards the data to the Data

Analytics module.

- A2D (analysis): The Data Analytics module performs data analytics on the received data changes, and generates and sends the data insights to the Decision module.
- D2E (decision): The decision module makes the decision on how to modify the existing services based on the data insights and calls MDSC (or PNC) to make service modifications.
- E2M (execution): MDSC (or PNC) executes the service modification by instructing the PNC (or NEs) to apply the service modifications or new network configurations. Optionally, MDSC (or PNC) may send feedback to the Data Monitoring module to modify its monitoring policy.

The middle block in Figure 4 represents the Knowledge functional block, which is a means for storing and retrieving data that is shared between the stages within a closed-loop. IETF's network management datastore architecture (NMDA) [14] is adequate to serve this purpose, and ACTN's YANG datastore complies with the requirements of NMDA.

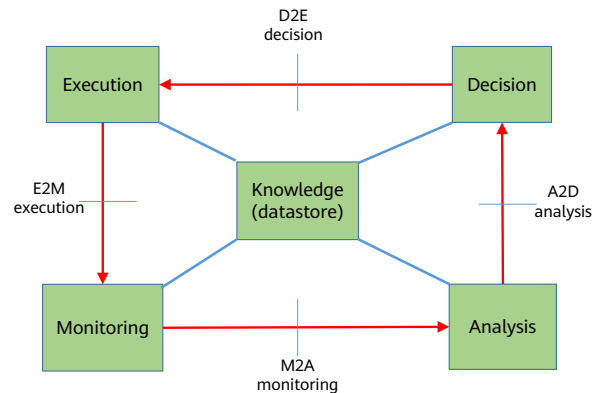


Figure 4 Closed-loop control flow defined by ETSI ZSM 009-1

2.3 Extended ACTN Architecture for Network Automation

In the previous sections, we showed that ACTN is an SDN architecture which provides network programmability and service orchestration over a multi-domain, heterogeneous network. Furthermore, ACTN can be extended to incorporate the IBN concepts in order to overcome some of the limitations of SDN and become a more capable solution for network automation. Such an extended architecture is shown in Figure 5. The extension mainly contains two parts. First, an Intent Translation functional block is added to the MDSC to provide the intent interface

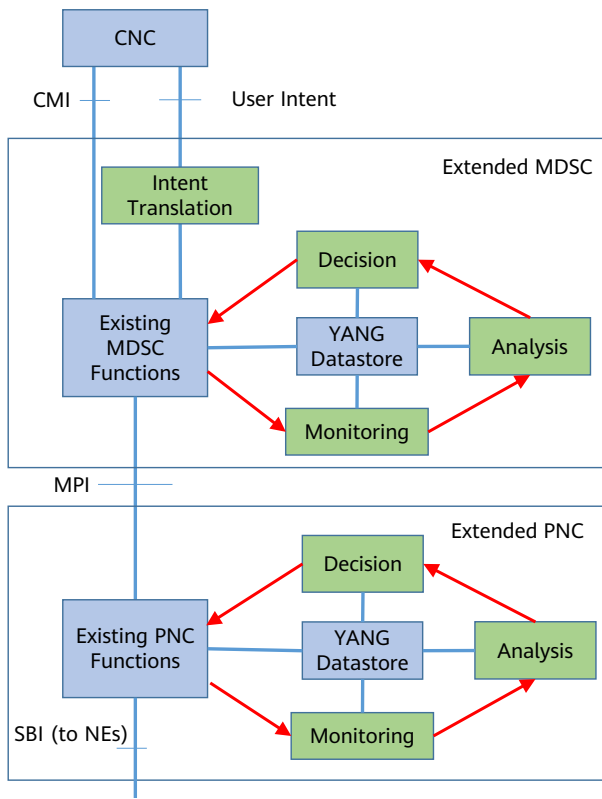


Figure 5 Extended ACTN architecture to support network automation

to the CNC. For backward compatibility, the existing CMI is intact. Although the CMI models are already intent-like and technology agnostic, the user intent will be even more high level and simple to use. In some use cases, user intent may be expressed using natural language, in which case the Intent Translation may need to support Natural Language Processing (NLP) [15].

The second part of the extension is the closed-loop mechanism. Due to the recursive nature of the hierarchical controllers, the closed-loop functional blocks in the MDSC and PNC, including Monitoring, Analysis, and Decision, are

very similar, implying that the common code infrastructure can be reused. In addition, these functional blocks may not be pure computer programming-based functional blocks. ML technologies are becoming more and more effective in this area [16]. Also note that the NMDA YANG datastore is used as the Knowledge functional block in closed-loop.

3 Innovations in All-Optical Networks and ACTN-Enabled Use Cases

This section discusses the technological innovations in all-optical networks and ACTN-based use case solutions, which may serve as enablers of the computing and networking convergence.

3.1 Innovations in All-Optical Networks

As mentioned earlier, future networks will need to be ultra-reliable, low latency (URLLC), high throughput, and energy and cost efficient. The recent developments in optical networks have made it well positioned to meet these requirements. Optical Transport Network (OTN) services are already known to be highly reliable and have guaranteed bandwidth and latency. Traditionally OTN services are only deployed in the core networks. In other words, OTN is usually used as an underlay service bearer rather than an end-to-end user service. However, the recent technology development is moving towards achieving ubiquitous all-optical connectivity, where OTN services are extended from the core to the edge, allowing it to directly provide

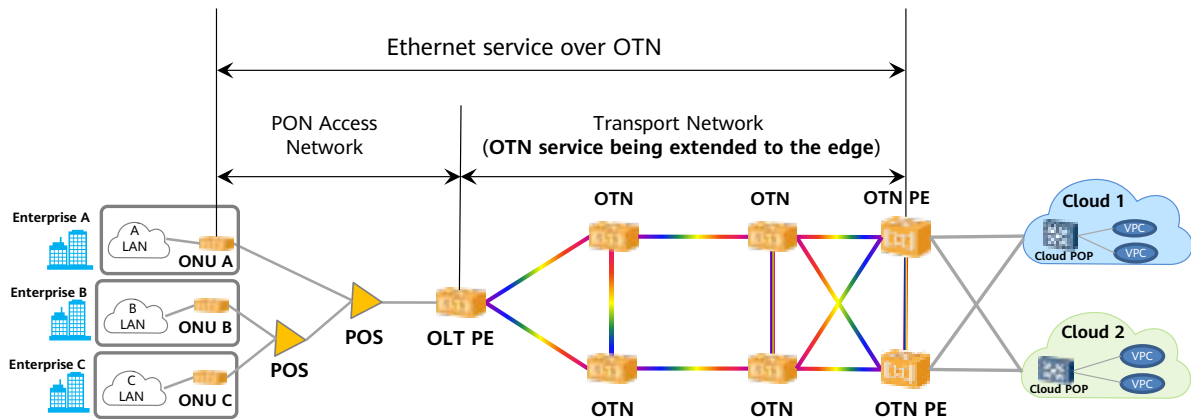


Figure 6 An all-optical network example which integrates OTN and PON technology to provide high-quality and cost-effective services

connectivity between end users and thus deliver end-to-end services. For example, in Huawei's access and transport network convergence solution [17], the optical line terminal (OLT) can function as an OTN provider edge (PE) node, so that the high-quality OTN services can be deployed at the edge. Furthermore, all-optical networks combine OTN and passive optical networks (PON) to gain benefits from both technologies. PON is a cost-effective solution which enables shared fiber connections for residential homes and enterprises. Thus, the resulting all-optical network can offer high-quality and cost-effective services. Figure 6 illustrates an all-optical network example.

However, one problem arises when moving OTN to the edge. The services in the edge network require a much larger service bandwidth span than that in the core. Therefore, although the current optical channel data unit - k (ODUK) multiplexing may be sufficient to achieve good bandwidth utilization in core networks, it is not suitable for edge networks. A solution to this problem is the small granularity switching with optical service unit (OSU) containers [18]. The OSU technology has the following characteristics.

- The OSU container can carry services with a minimum bit rate of 2 Mbit/s, which is 500 times smaller than that of the traditional smallest OTN container, which is 1.25 Gbit/s (ODU0). Moreover, as OSU is compatible with smaller service bandwidth, one optical link can carry 1000 OSU connections, whereas the same link can only support up to 80 traditional OTN connections.
- OSU is service-oriented. An OSU container can be identified with an attached tributary port number (TPN), which can be used to identify the user service carried by the container. In other words, there is a mapping relationship between the OSU and its corresponding service.
- OSU is backward compatible with the existing OTN architecture. An OSU can either be multiplexed into

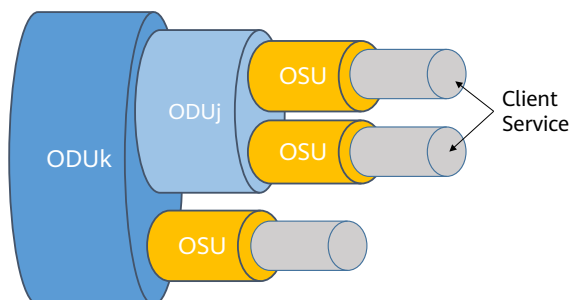


Figure 7 OSU multiplexing, compatible with the existing OTN architecture

lower order ODUj and then multiplexed into higher order ODUk, or be directly multiplexed into higher order ODUk, as shown in Figure 7.

In summary, an all-optical network, which combines the highly flexible OSU-enabled OTN technology and the cost-effective PON technology, is well positioned to meet the requirements of future networks.

3.2 All-Optical Slicing Use Case

Although many use cases and applications can be developed on the all-optical network, we discuss two particular use cases which are relevant to the integrated networking and computing platform discussed in the next section. The first relevant use case is all-optical slicing [10], which enables virtualization of the all-optical network. In other words, it enables a network operator to offer differentiated services to multiple tenants using a single network. A network slice is a logical network which has the following characteristics [19].

- It has its own topology structure (which could be modeled by using the ACTN VN model [8]) and owns a set of network resources which are used to provide connectivity. In addition, it also defines a Service Level Agreement (SLA) with which the connection needs to comply.
- It is isolated from other network slices. Each network slice is managed and controlled independently, so that changes to one slice do not impact other slices.

An all-optical network can create network slices by configuring OTN resources and services [20]. An OTN slice can be link-based, in which an entire OTN link is allocated exclusively to it. Alternatively, it can also be tributary-slot based, in which it shares an OTN link with other slices by allocating certain tributary slots on that link. Note that the former is a special case of the latter, because the former is achieved by allocating all the tributary slots to an OTN slice.

An OTN slice provides both resource and service isolation. Figure 8 shows an example of OTN slicing, in which two slices are created. The resources of the two slices are isolated by partitioning the tributary slots in the OTN TE link. Connectivity on a slice is provided by OTN tunnels which consume the allocated tributary slots within the slice. The source and the destination of an OTN tunnel are Tunnel Termination Points (TTP), as shown in Figure 8. This example shows an Ethernet service (e.g., EPL or EVPL) as

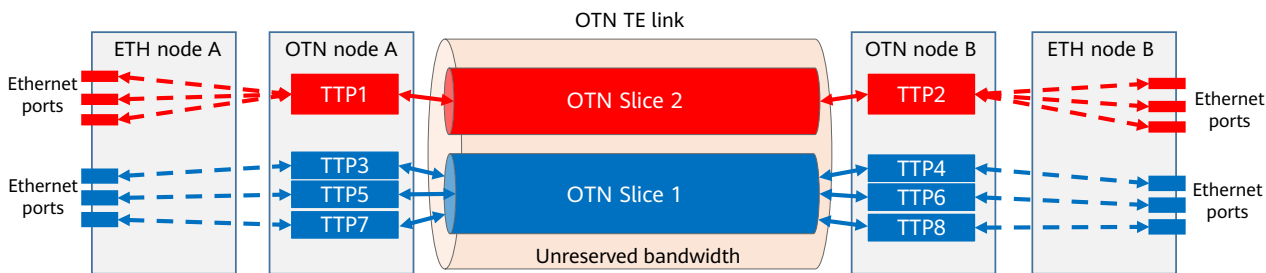


Figure 8 An Example of tributary-slot-based OTN slicing which carries Ethernet services

the end-to-end client service deployed on the slices. Service isolation is achieved by mapping the Ethernet services to their corresponding OTN tunnels in different slices.

The applicability of ACTN to network slicing is discussed in [21]. A customer's CNC may use the ACTN VN model to define a network slice and pass it to the MDSC, via the CMI. The MDSC then translates the network slice instance (VN) into the actual OTN slicing configurations and sends them to the PNCs, via the MPI. Then PNC configures the OTN services on the physical network.

3.3 Cloud-Optical Integration

Cloud-optical integration enables network operators to offer services that allow enterprises, governments, and home residences to access cloud computing services in a ubiquitous way. Figure 9 depicts a simplified view of the use case. The all-optical network shown in the figure can use Ethernet-over-OTN (EEO) technology to connect enterprise networks with clouds. User data (e.g., IP packets or Ethernet frames) are carried by Carrier Ethernet services, which operate over Ethernet Virtual Connections (EVCs). The EVCs provide service Operations, Administration, and Maintenance (OAM), whereas service isolation and traffic protection are offered by the underlay OTN services. Right-sizing and topological configuration of service tunnels, or groups of tunnels, is provided by reconfigurable OSUflex and DWDM technologies.

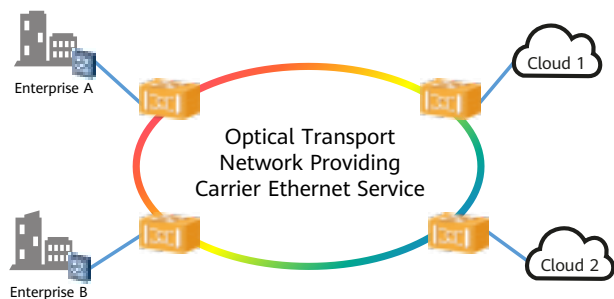


Figure 9 Optical network connecting enterprises and cloud

Note that the all-optical slicing solution can be used in this use case to provide service isolation among the tenants. In terms of Ethernet service models, a P2MP Ethernet service model, such as Ethernet-Tree (E-Tree), needs to be developed in order to effectively support the cloud private lines between enterprises and clouds. Figure 10a depicts an ACTN architecture for the cloud private line use case. The customer's CNC can use a VN model to define the requirement for the cloud private line and send it to the MDSC. The MDSC maps the VN model to the E-Tree service model and sends it to the converged optical controller (i.e., PNC), which delivers the E-Tree service on the optical network.

Figure 10b depicts the E-Tree model. The OTN PE connecting to the cloud POP is the E-Tree root. Primary and secondary E-Tree roots are supported for protection (e.g., Cloud 1). The OLT PEs connecting the enterprise networks serve as E-Tree leaves.

4 ACTN-based Infrastructure for Network, Computing, Storage and User Data Integration

As mentioned earlier, another essential requirement for future networks is to facilitate ubiquitous computing, i.e., making computing power a pervasive resource, just like electricity [22]. This requires the integration of computing and networking, and research in this field has started to attract attention from the industry, through standards organizations as well as in the academia. Possible solutions, such as Compute First Networking (CFN) [23], have been proposed. CFN is a recent advancement of the Information Centric Networking theory [24]. The idea is to provide a general-purpose computing platform by constructing a distributed compute graph consisting of compute nodes. A compute node is able to execute a computer program, join

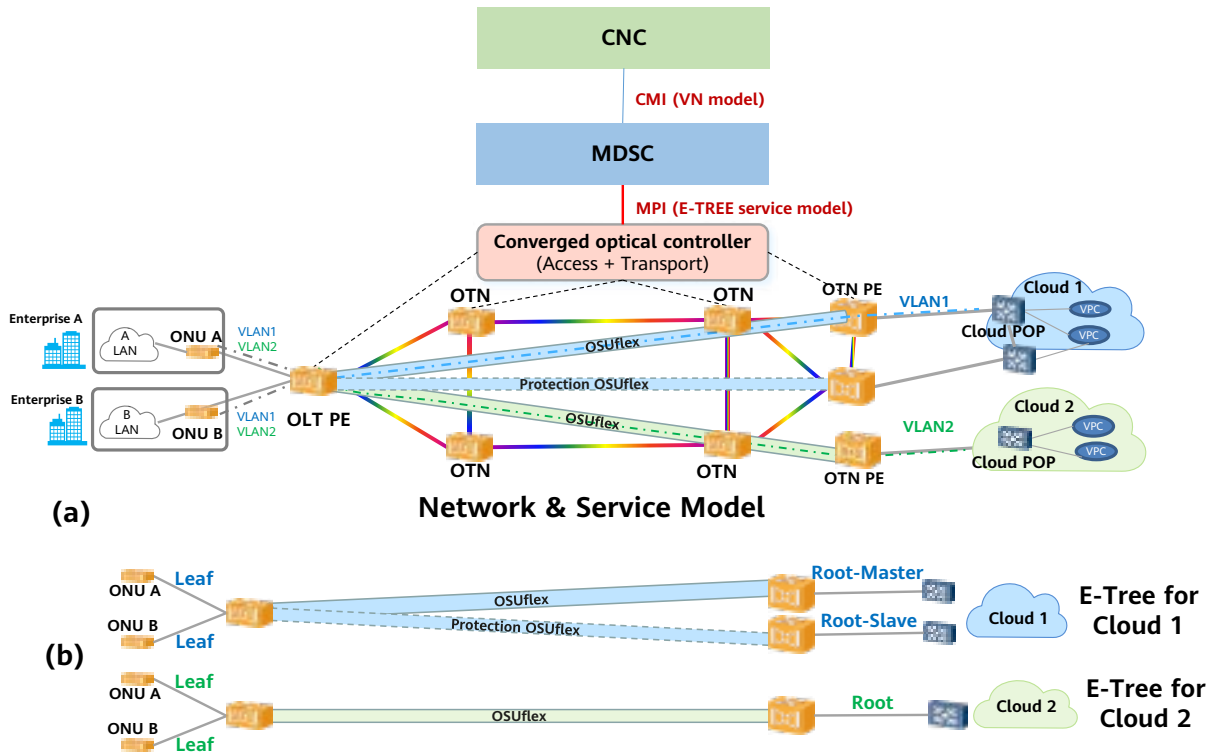


Figure 10 ACTN architecture and E-Tree service model for the cloud private line service

other nodes in the compute graph to form a multi-node execution, and schedule a task for future execution. The realization of the compute graph may need enablers, such as In-Network Computing [25], where network elements perform both networking and computing functions. However, In-Network Computing is a revolutionary networking paradigm which is in its early research stage, and solutions need to be found for many challenges and issues, such as data security, computing source discovery and allocation, and integrated service assurance and optimization, before this technology can be widely deployed.

Therefore, an evolutionary path to computing and networking integration is necessary; such a path is proposed

in [26]. The first stage of the suggested evolution is a cloud-network integration solution, where data centers in the clouds provide computing, storage and data-lake resources, and all-optical network solutions provide networking. The advantage of this solution is two-fold. First, most of the existing investment in cloud and network infrastructure can be reused by this solution. (In contrast, In-Network Computing may require a new generation of network equipment to be available.) Secondly, this solution is technology-ready for implementation. The enablers of the solution have been identified and discussed in the previous sections. Figure 11 shows these enablers and their relations.

Before we present the solution, it is worth noting that computing and networking convergence will create a paradigm shift that will result in profound changes to the business models of network operators and over-the-top (OTT) service companies. One big impact would be in the user data management area. In the current business model, user data is created, managed, and operated by the OTT companies (e.g., Google, Facebook, and Amazon). Moreover, user data and the OTT application (or service) are tightly coupled, causing user lock-in. For example, a user's account in Facebook and his/her social network information are locked-in with Facebook, making it difficult for the user to switch to other social media platforms. The new paradigm, however, may fundamentally change this by allowing the

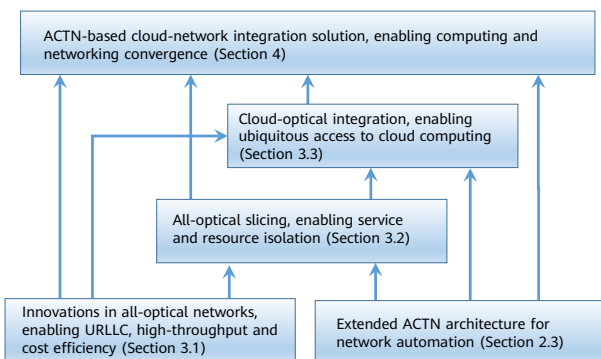


Figure 11 Technology enablers of the proposed cloud-network integration solution

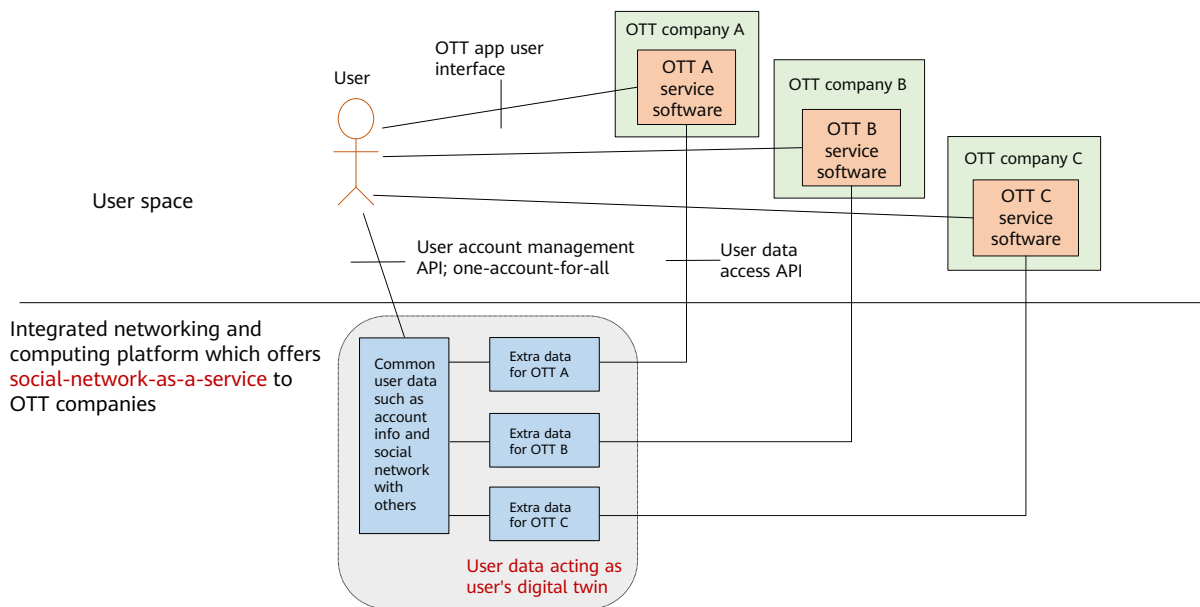


Figure 12 Social-network-as-a-service which decouples user data from OTT applications

network operators to offer social-network-as-a-service, which decouples user data and its operations from the OTT services that use these data. Figure 12 illustrates this idea.

In this new paradigm, the user only needs to create one user account regardless of the number of OTT applications that they want to use. This 1:1 mapping relationship between the user data and the user may then allow the data object to act as a digital twin of the user. For example, the user data object may reference data objects of other people to form a relation graph which mirrors the user's social relations in real life. This relation graph (i.e., social network), however, is stored and maintained by the platform, rather than by the OTT applications. OTT applications can retrieve or update the social network information via the platform's APIs. (This call flow is similar to ACTN's MDSC retrieving or updating the abstract network topology from the PNC.) This decoupling makes it easier for the user to switch between OTT applications.

Also shown in Figure 12, the user data can be accessed and managed directly by the user via the platform's northbound interface. For example, when the user wants to subscribe to or unsubscribe from an OTT service, instead of creating or deleting an OTT account, the user can grant or remove the permission of the OTT's access to the user's data.

The new platform integrates and orchestrates computing, networking, storage and user data resources to provide a unified service. Figure 13 depicts the use case of the platform. The platform provides common computing and network facilities to OTT services. The end user directly

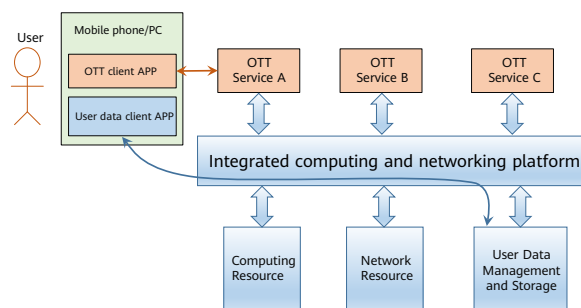


Figure 13 Integrated computing and networking platform use case diagram

interacts with the OTT services in the same way as today, and can also directly access and manage their data.

In the new integrated computing and networking paradigm, the meaning of network is extended. A network is no longer perceived as mere connections among the network elements, but, rather, a connected graph of the user data objects, which model the end users (i.e., users' digital twins). The value of the new platform is to offer the management of these connected graphs of user objects as an independent and common service which can be consumed by the OTT services. Furthermore, the platform needs to meet the following design requirements.

- The interface between OTT applications and the platform should be intent-based, allowing the OTT applications to specify the business needs without considering how to connect the underlay networks.
- The platform needs to map (or translate) the intent of the OTT to a collection of computing, network, and user

data resources, and allocate these resources accordingly.

- The platform manages the user data objects and their relation graphs independently, decoupled from OTT services. However, OTT services can read from and write to the user data objects according to the permissions set by the end users.
- OTT services are isolated from each other.

An ACTN-based solution to this platform is presented in Figure 14. The solution contains three layers. The top layer presumes the functionality of the CNC. It provides integrated computing and networking services to OTT businesses. The interface between the CNC and OTT is intent-based, allowing the OTT to define, retrieve, compute, and update a social network consisting of user data objects without knowledge of the underlay network technology.

The CNC provides end-to-end service management and orchestration. It translates and decomposes an OTT's service intent to the corresponding computing, network, and user data resources and services. It sends the network portion of the decomposed intent to the MDSC. The interface between the CNC and the MDSC is also intent-based, in order to reduce the complexity of the CNC. In addition, the CNC also interacts with the data centers for user data management operations.

The MDSC and the PNCs are responsible for carrying out the network operations including all-optical slicing and cloud private lines. The extended ACTN architecture (presented in section 2) is used, so that the MDSC and the PNCs have the Intent translation and analytics-based closed-loop capabilities.

5 Future Work

Although innovations in all-optical networking technologies and ACTN-based network automation solutions are promising in terms of meeting the requirements of future networks, there is much research to be done in this area. In the networking protocols area, an effective solution for establishing MP2MP connections over OTN needs to be researched [27]. Such a solution may require the OTN edge equipment (OTN PE) to have packet processing functions, e.g., IP-aware OTN switching.

In the management and control area, both the research and the standardization of IBN are in progress. The Intent data model and Intent life cycle management, for example, will eventually become standardized. The ACTN standards should incorporate or comply with the Intent standards in the future.

In the network automation area, AI/ML-based closed-loop mechanism is also worthy of research. Recent researches have shown some good results from ML-based data analytics and policy generation algorithms in optical networks [16].

6 Conclusions

This paper reviews the state-of-the-art innovations in optical networks in the context of meeting the requirements of future networks. Although many challenges and issues may still exist and further research is required, these innovations

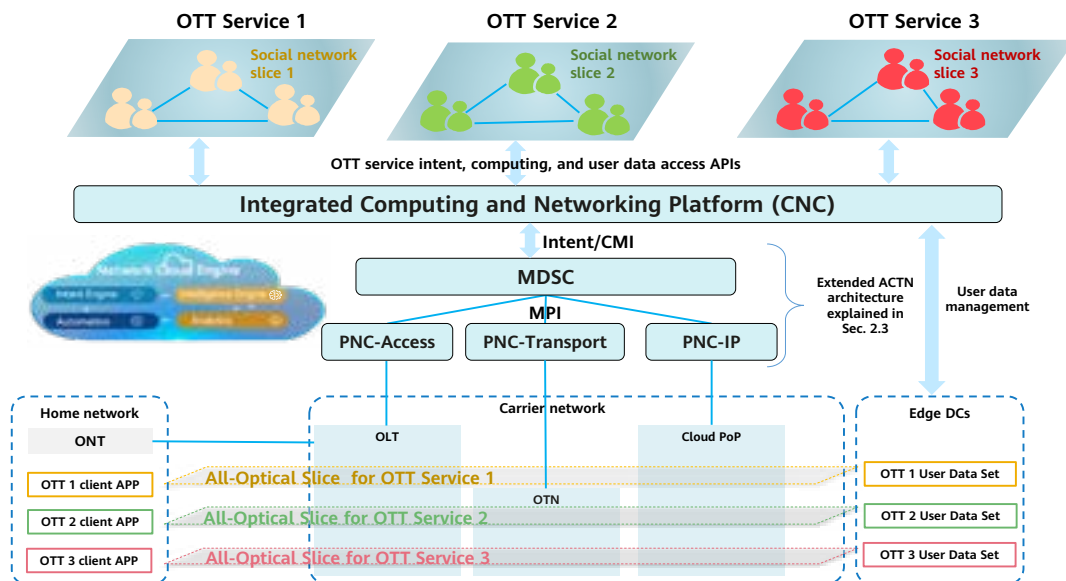


Figure 14 ATCN-based solution to integrated computing and networking platform which offers social-network-as-a-service

are nevertheless important steps in the evolution of optical networks. In addition, we show that the ACTN framework, in conjunction with IBN, may be an enabler for network automation. We further show that ACTN, being a flexible and scalable architectural solution, can be used to build an integrated computing and networking service platform.

References

- [1] D. Soldani and A. Manzalini, "Horizon 2020 and beyond: on the 5G operating system for a true digital society," *IEEE Vehicular Technology Magazine*, vol. 10, no. 1, pp. 32-42, March 2015.
- [2] D'souza, Chris, and David Williams, "The digital economy," *Bank of Canada Review*, 2017.Spring (2017): 5-18.
- [3] N. Adem, A. Benfaid, R. Harib, and A. Alarbi., "How crucial is it for 6G networks to be autonomous?," *arXiv*, June 2021.
- [4] TMF, "Autonomous networks: empowering digital transformation for the telecoms industry," White Paper, [Online]. Available: <https://www.tmforum.org/wp-content/uploads/2019/05/22553-Autonomous-Networks-whitepaper.pdf>.
- [5] M. Giordani, M. Polese, M. Mezzavilla, S. Rangan, and M. Zorzi, "Toward 6G networks: use cases and technologies," *IEEE Communications Magazine*, vol. 58, no. 3, pp. 55-61, March 2020.
- [6] X. Tang *et al.*, "Computing power network: the architecture of convergence of computing and networking towards 6G requirement," *China Communications*, vol. 18, no. 2, pp. 175-185, Feb. 2021.
- [7] D. Ceccarelli and Y. Lee, "Framework for abstraction and control of traffic engineered networks (ACTN)," *IETF RFC8453*, [Online]. Available: <https://tools.ietf.org/html/rfc8453>.
- [8] Y. Lee, S. Belotti, D. Dhody, D. Ceccarelli, and B. Yoon, "Information model for abstraction and control of TE networks (ACTN)," *IETF RFC8454*, [Online]. Available: <https://tools.ietf.org/html/rfc8454>.
- [9] Boucadair, M. and C. Jacquenet, "Software-defined networking: a perspective from within a service provider environment," *RFC 7149*, [Online]. Available: <https://tools.ietf.org/html/rfc7149>.
- [10] Huawei, "ACTN 2.0 technical white paper,"

- White Paper, [Online]. Available: <https://e.huawei.com/en/material/networking/enterprise/28e84f8e324a4a7f97e2e315e55dcc6a>.
- [11] L. Velasco, S. Barzegar, F. Tabatabaeimehr, and M. Ruiz, "Intent-based networking and its application to optical networks [Invited Tutorial]," *J. Opt. Commun. Netw.* 14, A11-A22 (2022).
- [12] A. Clemm, L. Ciavaglia, L. Granville, and J. Tantsura, "Intent-based networking - concepts and definitions," *IRTF draft work-in-progress (2021)*.
- [13] ETSI, "GS ZSM 009-1 V.1.1.1: zero-touch network and service management (ZSM); closed-Loop automation; part 1: enablers," *ETSI Group Specification*, [Online]. Available: https://www.etsi.org/deliver/etsi_gs/ZSM/001_099/00901/01.01.01_60/gs_ZSM00901v010101p.pdf.
- [14] M. Bjorklund, J. Schoenwaelder, P. Shafer, K. Watsen, and R. Wilton, "Network management datastore architecture (NMDA)," *IETF RFC8342*, [Online]. Available: <https://tools.ietf.org/html/rfc8342>.
- [15] GLUE, "GLUE Benchmark Leaderboard," [Online]. Available: <https://gluebenchmark.com/leaderboard>.
- [16] L. Velasco *et al.*, "Autonomous and energy efficient lightpath operation based on digital subcarrier multiplexing," *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 9, pp. 2864-2877, Sept. 2021.
- [17] Huawei, "Edge OTN technical white paper," White Paper, [Online]. Available: <https://carrier.huawei.com/~media/CNBGV2/download/products/networks/edge-en.pdf>.
- [18] L. Bai, "Optical service unit (OSU)-based next generation optical transport network (NG OTN) technology and verification," in *2nd International Conference on Computer Science Communication and Network Security (CSCNS2020)*, MATEC Web Conf. vol. 336, 2021.
- [19] A. Farrel, *et al.*, "Framework for IETF Network Slices," *IETF Internet Draft*, [Online]. Available: <https://tools.ietf.org/html/draft-ietf-teas-ietf-network-slices>.
- [20] A. Guo, *et al.*, "Framework and data model for OTN network slicing," *IETF Internet Draft*, [Online]. Available: <https://tools.ietf.org/html/draft-ietf-ccamp-yang-otn-slicing>.
- [21] D. King, J. Drake, H. Zheng, and A. Farrel, "Applicability of abstraction and control of traffic engineered networks (ACTN) to network slicing," *IETF Internet Draft*, [Online]. Available: <https://tools.ietf.org/html/draft-ietf-teas-applicability-actn-slicing>.
- [22] China Academy of Information and Communications Technology (CAICT), "White paper on China's computing power development index," White Paper, [Online]. Available: <https://cset.georgetown.edu/publication/white-paper-on-chinas-computing-power-development-index/>.
- [23] M. Król, S. Mastorakis, D. Oran, and D. Kutscher, "Compute first networking: distributed computing meets ICN>," in *Proceedings of the 6th ACM Conference on Information-Centric Networking (ICN '19)*, Association for Computing Machinery, New York, NY, USA, 67-77, Sept. 2019.
- [24] G. Xylomenos *et al.*, "A survey of information-centric networking research," *IEEE Communications Surveys & Tutorials*, vol. 16, no. 2, pp. 1024-1049, Second Quarter 2014.
- [25] I. Kunze, *et al.*, "Use cases for in-network computing," *IRTF Internet Draft*, [Online]. Available: <https://tools.ietf.org/html/draft-irtf-coinrg-use-cases>.
- [26] X. Duan, H. Yao, Y. Fu, L. Lu, and T. Sun, "Computing force network technologies for computing and network integration evolution," *Telecommunications Science*, 2021, 37(10): 76-85.
- [27] S. Liu and H. Zheng, "Accessing cloud via optical network problem statement," *IETF Internet Draft*, [Online]. Available: <https://tools.ietf.org/html/draft-liu-rtgwg-optical2cloud-problem-statement/>.



Development and Application of Optical Fiber Sensing Technology

Hao Li ^{1,2}, Cunzheng Fan ^{1,2}, Xiangpeng Xiao ^{1,2}, Liuyang Yang ^{1,2}, Liangye Li ^{1,2}, Baoqiang Yan ^{1,2}, Junfeng Chen ^{1,2}, Yuejuan Lv ^{1,2}, Haoguang Liu ^{1,2}, Yuze Dai ^{1,2}, Zhijun Yan ^{1,2}, Qizhen Sun ^{1,2,3,*}

¹ School of Optical and Electronic Information at Huazhong University of Science and Technology

² National Engineering Research Center of Next Generation Internet Access System

³ Wuhan National Laboratory for Optoelectronics

Abstract

Sensing technology is an important modern information technology. Optical fiber sensing uses optical fiber as a means to relay sensing information. Combined with transducer structures and modification materials, optical fiber sensors can sense various physical, chemical, and biomass parameters such as the external temperature, pressure, strain, acceleration, acoustic wave, refractive index, and protein. Due to the unique advantages such as high sensitivity, anti-electromagnetic interference, light weight, small size, high adaptivity to extreme environments, and easy multiplexing and networking, optical fiber sensing has started to be applied in various fields and is gradually becoming a mature industry. Based on the characteristics of the sensor architecture, this paper provides an overview for point-type, quasi-distributed, and distributed optical fiber sensors, and details the basic principles, key technologies, system research, and application progress of typical devices and systems such as fiber grating, Fabry-Perot Interferometric sensors, spectral absorption sensors, and distributed sensors based on optical fiber scattering. In addition, it outlines the future development prospects and application directions of optical fiber sensing.

Keywords

optical fiber, optical fiber sensing technology, fiber grating, fiber interferometer, distributed fiber sensing

* Corresponding author.

1 Introduction

With the rapid development of the Internet of Things (IoT), information sensing is increasingly important in all aspects of daily life and production. Sensing technology is one of the most important and rapidly developing cutting-edge technologies. Sensing, communication, and computer technologies constitute the three pillars of the information industry in modern society. Optical fiber sensing is a new sensing technology that uses light as the carrier of information and uses optical fibers as the transmission and sensing media. Through analyzing optical signal changes in the sensing fiber, the useful information is obtained. Optical fiber sensing has many unique advantages, such as small size, light weight, resistance to corrosion and high temperature, anti-electromagnetic interference, large information capacity, and high sensitivity. In addition, it supports large-scale and long-distance signal detection by using multiplexing technologies or distributed structures. This makes optical fiber sensing especially suitable for long-term on-line measurement in harsh environments, and highlights its advantages over traditional electromagnetic sensing. After more than 40 years of academic research and technical development, optical fiber sensing technology has developed rapidly in recent years. There is extensive scientific research on the measurement of physical parameters such as temperature, pressure, sound waves, vibration, magnetic field, and strain. In addition, it has been widely used in multiple scenarios and fields, such as electric power, petroleum, chemical industry, transportation, biology, and medical care. The fiber sensing technology has gradually formed a pattern of integration of industry, academia, and research.

Based on the characteristics of sensor architecture and measurement scopes, this paper provides an overview for point, quasi-distributed, and distributed optical fiber sensors and details the basic principles, key technologies, system research, and application progress of fiber grating, Fabry-Perot Interferometric sensors, spectral absorption sensors, and distributed sensors based on optical fiber scattering. In the end,

2 Single-Point Fiber Sensor and Its Application

it outlines the future development of optical fiber sensing.

Single-point fiber sensors are generally used to obtain local

information, and mainly include fiber grating sensors, fiber microstructure interference sensors, fiber spectral absorption sensors, and micro-nano fiber sensors [1]. These sensing units can implement sensing and measurement of one or multiple physical parameters, such as stress, vibration, refractive index, and temperature. Recent years have seen significant advancements to the sensitivity of a single-point fiber sensor and the range of measurable physical

2.1 Fiber Grating Sensor and Multi-parameter Sensing Application

parameters through the use of various optical, material, and mechanical sensitization and transduction mechanisms.

A fiber grating is a diffraction grating formed by periodically modulating the refractive index along the axial direction of an optical fiber. It has a special coupling effect on light in a certain wavelength range. Due to the benefits such as high sensitivity, anti-electromagnetic interference, anti-corrosion, wide dynamic measurement range, small size, high stability, and easy multiplexing, fiber grating sensors are the most widely used sensors in recent years. Common fiber gratings for sensing include fiber Bragg grating (FBG) sensors, long

2.1.1 FBG Sensor

period fiber grating (LPFG) sensors, and tilted fiber Bragg grating (TFBG) sensors.

An FBG sensor has a refractive index period of hundreds of nanometers, and can couple forward propagation core modes to backward propagation core modes within a specific wavelength range. An FBG sensor features narrow bandwidth and high reflectivity in reflection spectrum. The expression of the center wavelength is as follows:

$$\lambda_B = 2n_{eff}A \quad (1)$$

The preceding formula is the grating equation of an FBG sensor, where n_{eff} is the effective refractive index of the fiber core, and A is the grating period. Multiple environment variables can cause wavelength deviation or changes in the maximum power at the center wavelength of an FBG sensor [2, 3]. The basic sensing formula is as follows:

$$\delta\lambda_B = 2n_{eff}\delta A + 2A\delta n_{eff} \quad (2)$$

Where δA is the variation of the grating period, and δn_{eff} is the variation of the effective refractive index of the fiber

core. The experimental results show that the strain and temperature sensitivity of an FBG sensor near the 1550 nm wavelength are about 1.2 pm/ $\mu\epsilon$ and 10 pm/ $^{\circ}\text{C}$. FBG sensors have been widely used in the field of structural health monitoring [4]. After using special packaging for protection or sensitization (as shown in Figure 1), FBG sensors can

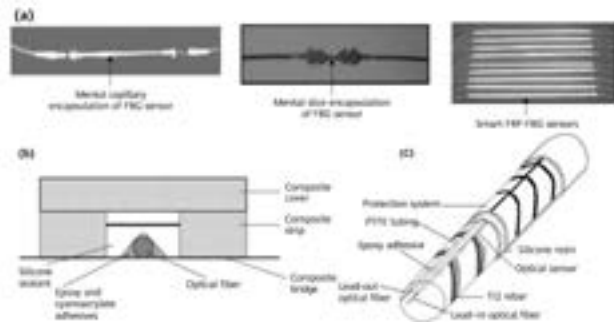


Figure 1 (a) FBG sensor packaging; (b) Bridge detection; (c) Pile load monitoring [6]

2.1.2 LPG Sensor

measure the internal strain and temperature of structures in a high resolution and large range.

Compared with an FBG sensor, a long period grating (LPG) sensor has a larger refractive index period, resulting in coupling of the forward propagating core mode to forward propagating cladding modes within a specific wavelength range [5]. The resonant wavelength expression of the cladding mode is as follows:

$$\lambda_L = (n_{eff}^{core} - n_{eff}^{clad})\Lambda \quad (3)$$

$\delta\lambda_L$ is the shift of the LPG resonant wavelength, n_{eff}^{core} and n_{eff}^{clad} are the effective refractive indexes of the fiber core and cladding, respectively. A loss peak with wide bandwidth can be observed in the transmission spectrum. The central wavelength is affected by the external environment and is very sensitive to physical parameters such as temperature, refractive index, and strain.

In practice, the low-order cladding mode of an LPG sensor is typically used for refractive index sensing. When the refractive index to be measured is lower than or equal to the refractive index of the fiber cladding, the sensitivity of the blue shift of the resonant peak wavelength can reach thousands of nm/RIU [6]. If the refractive index of a medium is higher than that of the fiber cladding, the LPG cladding must be coated using a sensitive film with a refractive index that varies with the external environment. In addition, long-period surface biological functionalization

can also be used for biochemical sensing to detect specific viruses and bacterial cells [7, 8], as shown in Figure 2.

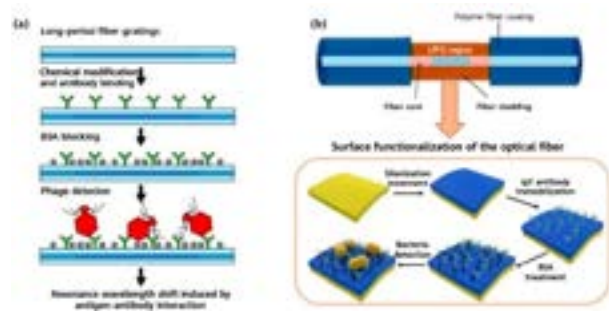


Figure 2 LPFG (a) Virus detection [7]; (b) Bacteria detection [8]

2.1.3 TFG Sensor

A TFG sensor couples the fiber core mode with the cladding mode by introducing a tilt angle in the fiber Bragg grating. As shown in Figure 3, tilted fiber gratings (TFGs) can be classified into tilted FBGs (TFBGs) and excessively tilted fiber gratings (ex-TFGs) depending on their tilt range. A TFBG couples the light that is transmitted in the forward fiber core mode within a specific wavelength range to the cladding mode in the opposite direction, while an ex-TFG couples the light to the forward propagation cladding mode. TFG sensors are widely used in biochemical sensing of hemoglobin, cells, and gas molecules due to its high-precision detection of temperature and refractive index [9].

The surface plasmon resonance (SPR) wave can be excited outside the cladding of a TFBG sensor by metal coating, enabling the refractive index to be measured with higher sensitivity. For different detection objects, surface biological functionalization can achieve higher sensitivity and specific biological detection, such as human acute leukemia cell line differentiation detection [10]. As shown in figure 3, if a magnetic fluid (Fe_3O_4) material is coated on the surface of an SPR-TFBG sensor, the direction of the magnetic field can be accurately measured by using the TFBG-specific "two-pole convergence" SPR energy field distribution feature. By measuring the directional scattering between the SPR strong resonance field and the magnetic nanoparticles under different magnetic fields, strong directivity measurement of the magnetic field intensity can be achieved [11].

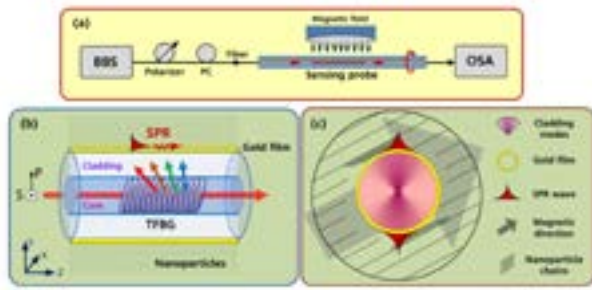


Figure 3 SPR-TFBG based magnetic field sensor [11]

2.2 Fabry-Perot Cavity Sensors and Pressure/Ultrasonic Detection Applications

As an important type of fiber sensors, fiber Fabry-Perot interferometric (FPI) sensors are widely used in pressure, temperature, and acoustic wave detection due to their high sensitivity, small size, and simple structure.

According to fiber functions in the sensor structure, the FPI sensors can be divided into intrinsic FPI (IFPI) sensors [12, 13] and extrinsic FPI (EFPI) sensors [14, 15]. Figure 4a shows the IFPI-type fiber sensor, which consists of two reflective surfaces inside a fiber. The fiber serves as both the light guiding component and the sensing component. An EFPI fiber sensor consists of sensing components attached to fibers. Generally, an EFPI fiber sensor consists of the end faces of two fibers or a fiber end face and a thin film, as shown in Figure 4b and Figure 4c, respectively. When parameters such as external pressure, sound waves, and temperature are applied to a fiber FPI sensor, structural deformation occurs, resulting in a distance or medium refractive index change between reflective surfaces. As a result, the multi-beam interference reflection spectrum or a transmission spectrum

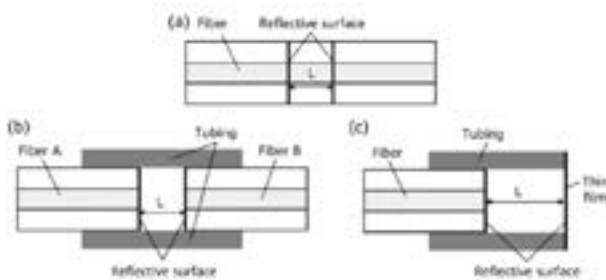


Figure 4 Schematic structural diagram: (a) Intrinsic fiber FPI sensor; (b) Traditional extrinsic FPI sensor; (c) Diaphragm type extrinsic FPI sensor

changes. A demodulation apparatus can be used to monitor the spectrum change of the FPI, thereby obtaining certain parameters to be measured.

Fiber FPI sensors are very sensitive to the changes of external pressure. Because the pressure sensitivity is closely related to the thickness and area of the fiber end face film, the pressure sensitivity can be significantly improved by increasing the area and reducing the thickness of the sensing film. As early as 2012, Zhixia Zheng et al. developed a fiber FPI pressure sensor using micro-electro-mechanical systems (MEMS) technologies. The sensor consists of a 6 μm thick sensitive film and can monitor the internal pressure of the human oral denture structure through power demodulation with a precision as high as 62 Pa [16]. Later, silicon dioxide films [17] and crystalline silicon films were used [18] to further improve the sensitivity, and MEMS technologies were used to precisely control the film thickness to ensure the consistency of sensors. Fiber FPI pressure sensors are also used to monitor human biological information, such as blood pressure. Fiso has developed a pressure sensor with a diameter of only 125 μm and a sensitivity up to 5 nm/kPa using MEMS technologies. The pressure sensor can be implanted in an infusion needle to measure the blood pressure with a precision as high as ±3 mmHg [19]. Fiber sensors processed by the laser hot melt technology can be used in harsh environments including high temperature and high pressure. Such sensors show good thermal stability at a temperature as high as 800°C, and can be used in long-term pressure monitoring for oil wells, engines, and gas turbines [20].

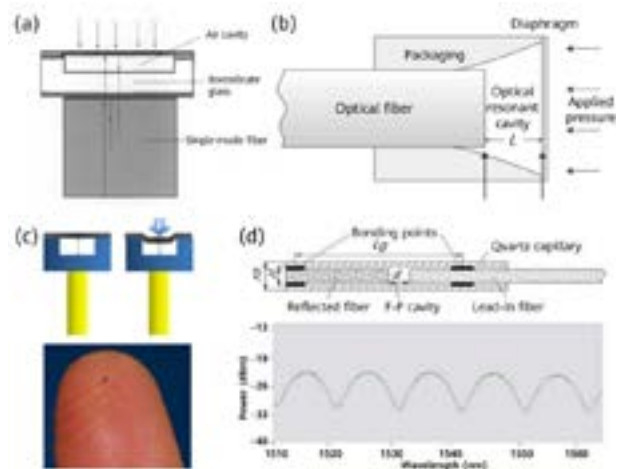


Figure 5 Fiber FPI pressure sensor: (a) Pressure sensor based on borosilicate glass [16]; (b) Pressure sensor based on a crystalline silicon thin film [18]; (c) Fiso pressure sensor for human blood pressure monitoring [19]; (d) High-temperature-resistant pressure sensor [20]

Ultrasonic waves are sound waves with a frequency greater than 20 kHz. Such waves feature short wavelength, strong directivity, and high resolution, and play an important role in industrial nondestructive testing, medical imaging, and partial discharge. Many types of fiber ultrasonic sensors have been developed, of which fiber FPI sensors have become popular in ultrasonic detection due to their excellent stability, high sensitivity, and compact structure [21–24]. As early as 1990, Alcoz et al. introduced TiO₂ film coating into the continuous single-mode fiber to form an FPI sensor structure. As shown in Figure 6a, this type of sensor is embedded in plastic and graphite composite materials to implement ultrasonic detection in the frequency band of 0.1 MHz to 5 MHz [25]. To further improve the sensitivity, a high-precision FPI-structure [26] can be obtained by improving the reflectivity of the end face, as shown in Figure 6b. In addition, a flat concave cavity can be designed to perform wavefront matching on the light emitted from a fiber, thereby reducing the optical transmission loss [27], as shown in Figure 6c. Furthermore, to solve the problem of uneven frequency response of fiber FPI ultrasonic sensors, researchers use two high reflectivity mirrors on the fiber end face to construct a rigid FPI for avoiding mechanical vibration of the structure [25], as shown in Figure 6d. At present, ultrasonic sensors of this type can achieve a noise equivalent sound pressure as low as 2 Pa and a measurement bandwidth of –6 dB for frequencies higher than 22.5 MHz [28]. These performance indicators are highly competitive.

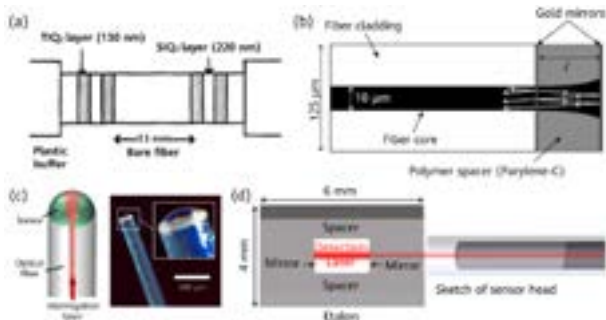


Figure 6 Fiber FPI ultrasonic sensor: (a) Intrinsic ultrasonic sensor [25]; (b) Ultrasonic sensor based on Parylene film [26]; (c) Wavefront matching ultrasonic sensor [27]; (d) Filmless ultrasonic sensor [28]

2.3 Spectral Absorption Sensor and Its Application in Gas Detection

A fiber spectral absorption sensor uses the absorption peak of the material to be measured in the transmission window of quartz fibers to measure the light intensity attenuation caused by the absorption of the material and obtain the material concentration. Assuming that the input light

intensity is I_0 , the relationship between the output light intensity I and gas concentration C can be described by the Beer-Lambert's law [29]:

$$I = I_0 \exp[-\alpha_0(\lambda)lC + \zeta(\lambda)] \quad (4)$$

In the formula, $\alpha_0(\lambda)$ is the near-infrared absorption coefficient of the gas to be measured, which is related to the wavelength. l is the length of interaction between the gas to be measured and light. $\zeta(\lambda)$ is the noise interference factor of the optical path. The detection based on the overtone band absorption spectrum in the near-infrared band can meet the requirements of gas detection in many fields. For this reason, fiber spectral absorption sensors are widely used. Compared with other types of fiber gas sensors, fiber spectral absorption sensors perform better in terms of high measurement sensitivity, high gas identification capability, fast response capability, simple and reliable gas sensor probes, and easy networking.

The absorption peak wavelengths of most gases are in the near-infrared band. The gas concentration can be obtained by measuring the light intensity change in the spectral absorption band of gas molecules in the fiber transmission window (near-infrared band). To eliminate the impact of the absorption coefficient and the noise interference of the optical path, the spectral absorption differential detection technology or the secondary harmonic detection method can be used. When two monochromatic light beams with very close wavelengths but a great difference in absorption coefficients pass through the gas to be measured at the same time or in a very short time, the relationship between the output light intensity of the two monochromatic light beams and the gas concentration can be obtained, and the concentration of the gas to be measured can be obtained by differential approximation [30]. Although the differential absorption detection can eliminate the interference of the optical path, the inherent noise of the system cannot be eliminated. Given this, the secondary harmonic detection method is generally used to extract the weak absorption signal [31]. As shown in Figure 7, a tunable narrowband laser is used to output sinusoidal frequency-modulation light. When the center wavelength of the light emitted by the laser is scanned near the gas absorption peak or is aligned with the gas absorption peak, the wavelength modulation is converted into light intensity modulation. The gas concentration can be obtained by analyzing the harmonic components of the light intensity. Using the ratio of the secondary harmonic to the primary harmonic as the system output, the common-mode noise such as light

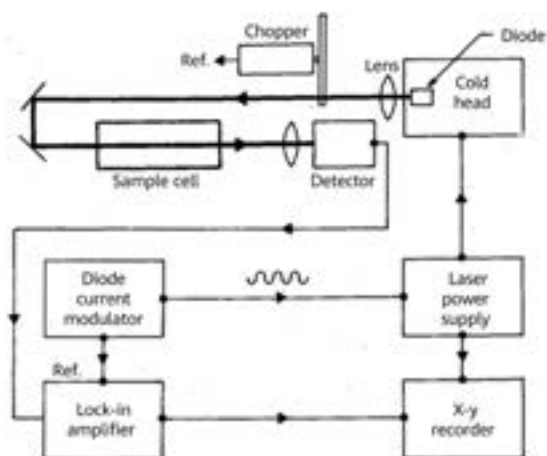


Figure 7 Schematic diagram of the secondary harmonic detection system [31]

source fluctuation can be further reduced, and the trace gas detection at the ppm level can be achieved.

The sensing unit of the spectral absorption gas sensor is crucial to the detection performance. The typical types include open absorption cavity and all-fiber cavity. In the open absorption cavity, the fiber is mainly used as the transmission medium. The light emitted from the fiber is collimated into parallel light using lenses. After being absorbed by the gas, the parallel light is converged into the output fiber by another group of lenses. In this way, the detection of multiple gases can be implemented with high accuracy and precision to the sub-ppb level [32]. The gas absorption of an all-fiber sensing unit occurs on the surface of a fiber or even inside a fiber after special processing. The fiber gas sensor based on the evanescent field can detect the surrounding gas by using the evanescent wave absorbed by the gas on the surface of a fiber. For a cladding-free fiber [33], we can further improve the sensing performance by adding a sensitive film that interacts with the gas to be measured. For a D-type fiber and micro-nano fiber [34], we can further improve the sensing performance by doping various nanomaterials in the cladding. As shown in Figure 8, a hollow core photonic band gap fiber (HC-PBF) is usually used as the sensing unit. The target gas is filled in the hollow area of the fiber in the gas chamber. The gas to be measured reacts with the laser of a specific wavelength in the hollow area. The gas concentration [35, 36] can then be detected by measuring the change of the transmitted light intensity. Because an all-fiber gas chamber may be more than tens of meters in length, sensing units of this type have higher detection sensitivity, and the lower limit of the detectable concentration is close to ppm.

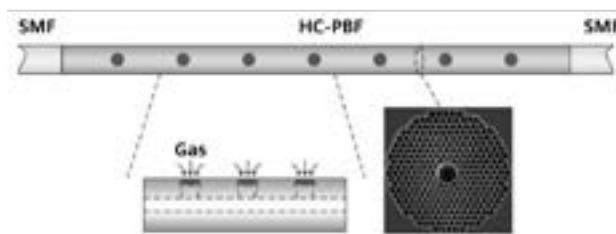


Figure 8 Fiber gas sensing unit based on HC-PBF [35]

2.4 Micro-Nano Fiber Sensor and Its Application in Health Monitoring

A micro-nano fiber (MNF) is a fiber with a diameter close to or smaller than the transmission wavelength. It has the characteristics of strong optical field constraint, strong evanescent field, surface field enhancement effect, and abnormal waveguide dispersion. Due to the direct contact and interaction between the large proportion of evanescent field on the surface and the external environment, an MNF is very sensitive to external physical parameters, and has an ultra-high refractive index sensitivity. It is widely used in the biochemical field, and can be used to quantitatively detect various viruses, cells, and microorganisms. With the development of wearable sensing technologies in recent years, MNF-based health monitoring has become a highly researched topic in the optical fiber sensing field.

2.4.1 Physiological Sensor Based on MNF

The real-time monitoring of one's vital signs, including heart rate, respiration rate, and blood pressure, can play a significant role in the early diagnosis and treatment of diseases. An MNF has a small size, high sensitivity, low bending loss, and high flexibility, and overcomes the disadvantages of complex fabrication processes and the high cost of flexible fibers. In recent years, multiple MNF-based wearable sensors, such as pulse sensors, respiratory sensors, and blood pressure sensors, have been successfully developed. It is, therefore, a promising option in the field of wearable devices.

In 2013, Nan-Kuang Chen et al. were able to successfully monitor resonance wavelength changes caused by vascular vibration signals using an erbium-ytterbium co-doped MNF mode interferometer [37]. In 2018, Jinhong Li et al. used a 100 nm-thick gold coating as a substrate, and placed an MNF junction resonant cavity into a PDMS thin film material to

create a sensor with excellent mechanical performance in the real-time measurement of wrist pulse signals [38]. In 2021, Heng-Tian Zhu proposed a stretchable ultra-thin optical sensor based on a self-assembled wave-shaped micro fiber, to collect human pulse signals in real time and dynamically and continuously monitor blood pressure parameters [39]. To make the sensor probe less demanding on the precision of position in the sensing area, Qizhen Sun et al. designed an MNF liquid hybrid optical chip, as shown in Figure 9. The chip supports high-fidelity pulse waveform recovery within 20 mm x 20 mm, and performs cardiovascular health assessments by extracting pulse wave pattern parameters [40].

Multifunctional skin sensors play an important role in next-generation healthcare, robotics, and bioelectronics. By embedding the MNF in a thin layer of polydimethylsiloxane (PDMS) in 2020, Lei Zhang et al. demonstrated a highly sensitive skin-like wearable optical sensor and successfully implemented pulse wave monitoring and human-machine interaction applications [41]. To improve sensitivity, the wave-shaped MNF is embedded in the PDMS to obtain a strain sensing coefficient of up to 750%, so as to monitor breathing and arm movement in real time [42]. In 2021, Qizhen Sun et al. proposed an optical neuron based on the MNF with flexible material packaging. The neuron

implements online quantitative perception of a finger bending action by tracking light intensity changes [43], and may be further used in the field of robotic arms and intelligent interactions.

2.4.2 Biochemical Sensor Based on MNF

Biochemical and physiological parameters are important indicators of human health. Since fibers are unable to directly identify biomasses, some functional materials are required to translate biochemical and physiological parameters into changes in physical parameters. These include refractive index, temperature, and stress, which can be directly identified by the MNF to indirectly sense different biochemical and physiological parameters. Based on the physical parameter sensors above and other advanced functional materials, researchers have further developed a variety of MNF biochemical sensors, including biochemical sensors that can sense DNA [44], proteins [45], biotoxins [46], cancer markers [47], gases [48], and heavy metal ions [49].

DNA carries unique and diverse biogenetic information. In 2011, M.I. Zibaii et al. achieved the on-line monitoring of DNA hybridization by detecting the spectral changes of non-adiabatic taper MNF [50]. In 2017, Yuanyuan Huang et al. reported an MNF-DNA sensor with graphene oxide coating. According to her team's paper, the sensor achieved a minimum detectable concentration of 1 pM by detecting spectral changes [51]. Proteins have many functions, such as replicating DNA, promoting the metabolism, responding to stimuli, and transporting molecules. In 2011, Zhang Lei et al. reported an MNF bovine serum protein sensor that can detect concentrations as low as 10 fg/ml. [45] In 2014, Kaiwei Li et al. applied micronano fibers to the detection of the cancer marker alpha fetoprotein [52], using two antibodies to form a sandwich structure with alpha fetoprotein to achieve selective detection of the target material. By using gold nanoparticles to amplify the evanescent wave absorption signals, the minimum detectable concentration can be reduced to 0.2 ng/ml, showing an extremely high detection sensitivity.

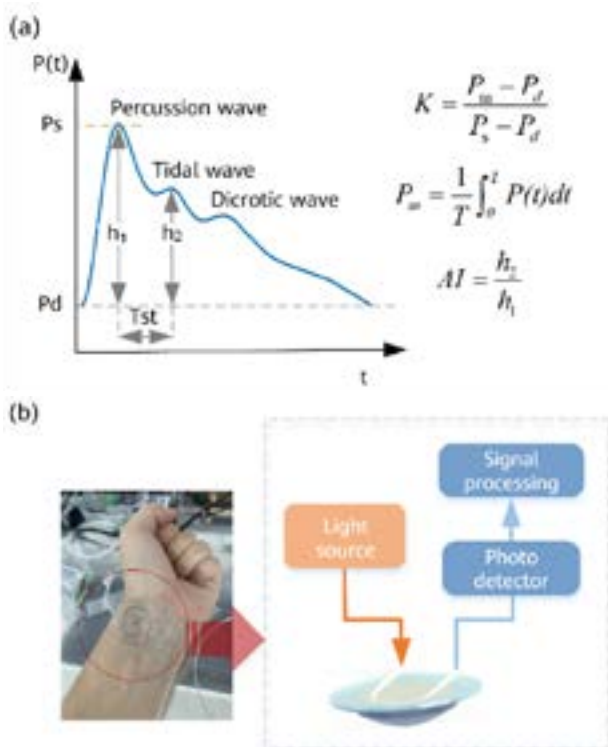


Figure 9 MNF-based calibration-free liquid hybrid optical chip [40]

3 Quasi-Distributed Fiber Sensor and Its Applications

Quasi-distributed fiber sensing technology uses a fiber to connect multiple sensing units in series to implement long-distance and large-capacity sensing and monitoring over a single fiber. Fiber Bragg grating, fiber FPI, and fiber interferometer are often used in quasi-distributed sensor networks. These sensor units have high signal-to-noise ratio (SNR) and reserve the advantages of distributed sensors in networking and signal transmission. A quasi-distributed sensor network can be formed on one or multiple fibers by using technologies such as wavelength division multiplexing (WDM), time division multiplexing (TDM), frequency division multiplexing (FDM), spatial division multiplexing (SDM), and hybrid multiplexing [53] based on the characteristics of optical waves (such as wavelength, time, encoding, and spatial channel attributes).

3.1 Quasi-Distributed Fiber Sensing Technology Based on the Fiber Grating Array and Its Typical Applications

3.1.1 Fiber Grating Array Fabrication Technologies

The capacity of a large-scale fiber grating array sensing network is not only related to the number of multiplexed gratings, but also related to the distance sensed by the network and the number of sensing parameters that can be monitored. Effective management and monitoring of each grating in large-scale fiber sensing networks will impose strict technical requirements on gratings and transmission fibers from optical and mechanical perspectives. Specifically, multiple technical parameters are involved, such as the transmission loss of a fiber, reflectivity and insertion loss of the grating, peak wavelength consistency of the grating, side mode suppression ratio of the grating, and spacing consistency of the grating. These indicators are closely related to the selection of fiber material and the grating fabrication process. The online fabrication of fiber grating arrays has become a mainstream technology, and is typically implemented using the following methods:

- Stripping and recoating: Strip the coating of a finished single-mode fiber manually or by using a CO₂ laser

[54], expose the fiber to an ultraviolet laser, engrave the FBG, and then coat the fiber again. This method is less demanding on the fabrication system, but is time-consuming and laborious. In addition, the mechanical strength of the fiber decreases after recoating.

- Fabrication based on a fiber drawing tower: As shown in Figure 10a, after fiber drawing and before coating, expose the fiber to a single pulse ultraviolet laser, engrave the FBG, and then perform the coating. In this way, a long-distance fiber grating array can be rapidly fabricated. The German FBGs sensor company [55] and the National Engineering Research Center for Fiber Sensing Technology of Wuhan University of Technology [56] have done much research on this technology. The disadvantage of this method is that it is difficult to control the consistency of parameters such as grating spacing, reflectivity, and peak wavelengths. The average cost of one drawing process is more suitable for mass production.
- Fabrication based on an ultraviolet-permeable coated fiber: Using a coated ultraviolet transparent fiber as the carrier and a high-precision continuous on-line grating engraving device based on ultraviolet laser exposure, fiber grating arrays of any length, spacing, and specifications can be rapidly fabricated. This method is highly flexible, consistent, and efficient, whose platform is shown in Figure 10b. Qizhen Sun from the School of Optical and Electronic Information at the Huazhong University of Science and Technology/National Engineering Research Center of Next Generation Internet Access System has the independent intellectual property rights of the entire chain — from coating materials and ultraviolet-permeable coated fiber fabrication, to grating array online fabrication technologies. A fiber

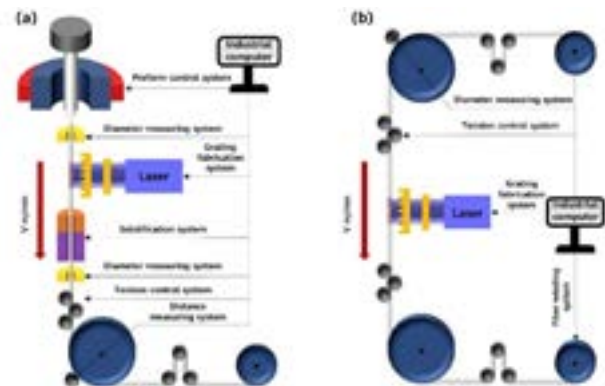


Figure 10 Schematic diagram of a large-scale online fabrication of fiber grating arrays: (a) Based on a fiber drawing tower; (b) Based on an ultraviolet-permeable coated fiber

microstructure array with a minimum spacing of 40 μm , more than 20 multiplexed wavelengths, and an unlimited single fiber distance has been successfully fabricated [57].

- Fabrication based on a femtosecond laser: Use the femtosecond laser to pass through the coating layer of the fiber to engrave the grating on the fiber core. The advantage of this method is that the grating can be flexibly engraved on any type of fiber (including the multi-core fiber), and the formed grating will not be erased at high temperatures. Sensors made using this method are suitable for sensing in extreme environments. However, due to the limitation of the processing technology, engraving speed is usually low and the insertion loss of the grating is high. Large-scale engineering applications based on this technology have not been reported yet [58].

3.1.2 Temperature Sensing Technology Based on Fiber Grating Array and Its Applications

As a temperature-sensitive component, a grating array sensor features electrical insulation and is especially suitable for scenarios such as long-distance fire alarms and large-scale temperature monitoring in places where electrical insulation is required. It has great promotion value in petroleum and petrochemicals, public railway tunnels, electric power, and green energy industries.

China's national standard GB16280-2014 for linear heat-sensing fire detectors formulated in 2014, requires that any sensitive components with a length of 100 mm must be able to quickly monitor temperature changes to generate early warnings for small fire sources. Currently, a temperature sensing cable can effectively sense such small fire sources, but sensing distance is short, and the fire source cannot be located. Therefore, it is unable to meet the requirements for long-distance fire detection and location. A mainstream Raman fiber temperature measurement system has a spatial resolution of about 1 m, and responds slowly to small fire sources or even fails to detect them. Based on hybrid TDM and WDM technologies, Qizhen Sun fabricated a weak grating array that integrates 600 sensor units with a spacing of only 10 cm on a single fiber based on ultraviolet-permeable coated fibers. Using a grating array demodulation method [53] based on a tunable fiber Fabry-Perot filter, the array supports quasi-distributed real-time

monitoring with a precision of 0.1°C , and has been used for thermal runaway monitoring of li-ion battery cell modules and energy storage power stations.

3.1.3 Strain Sensing Technology Based on Fiber Grating Array and Its Applications

Underground rail transit and large bridges will face various structural problems over time, caused by rock and soil pressure changes, the geological environment, and third-party construction. Therefore, there is an urgent need for online monitoring technology. As shown in Figure 11, Qizhen Sun designed a temperature self-compensated tunnel convergence monitoring sensor based on aluminum sheets. With temperature grating and strain grating connected in series and fixed in bending form to eliminate the impact of stress, the sensitivity of the strain grating of the sensor array can reach 8.28 pm/mm . Grating array sensing networks based on hybrid WDM and SDM have been successfully applied to the deformation monitoring of subway tunnel tube structures.

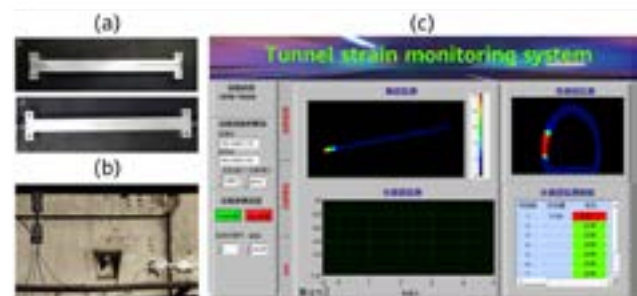


Figure 11 On-line deformation monitoring of subway tunnel tube structures: (a) Pictures of a sensor; (b) Onsite installation diagram; (c) Monitoring and warning system

3.2 Quasi-Distributed Fiber Sensing Technology Based on Interferometric Structure Array and Its Application in Underwater Detection

The structure of an interferometric fiber sensor unit mainly includes Michelson, Mach-Zehnder, Fabry-Perot, and Sagnac interferometers. Generally, the TDM technology is used for array networking to construct a quasi-distributed sensor system, which is typically applied to hydrophone arrays [59]. In 1977, the U.S. Naval Research Laboratory (NRL) first proposed the concept of fiber hydrophones [60]. Since then,

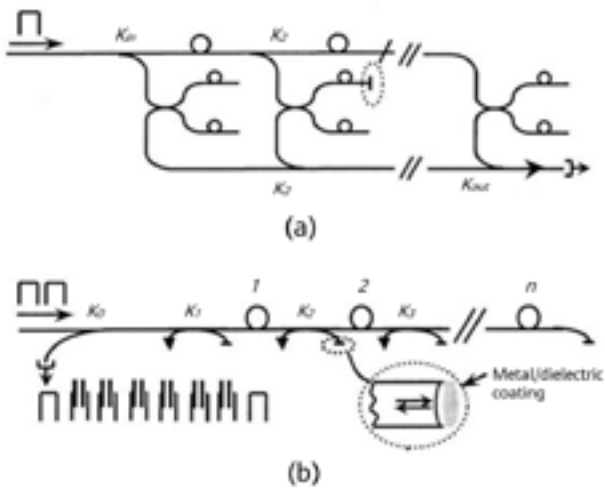


Figure 12 Structures of interferometric arrays [61]: (a) Bus-interferometer-bus structure; (b) Trapezoidal structure

large-scale networking technologies for interferometric fiber sensing have been highly valued, but there are few reports on them due to military restrictions.

Typical structures of interferometer arrays mainly include the following [61]:

- bus-interferometer-bus structure: The detection system required by an array in this structure is simple, but the process of the array configuration is complex.
- Near-end interferometric single-branch array: In this structure, each element of the array no longer has a conventional interferometer, but has one branch that is sensitive to sound signals. To enable an array of this type to generate interference signals normally, a special modulation or demodulation mode is required at the transmit end or receive end, so as to effectively reduce system noise. In addition, as the main line is sensitive to external signals, it requires special vibration isolation.
- Trapezoidal structure: This kind of array is simpler in structure as the symmetric part is omitted from the

second array by way of reflector, and only one bus is used to complete the distribution and recovery of energy. In addition, because a Faraday rotator is used at the reflector end, the polarization fading problem is resolved. With the improvement of fiber coupler fabrication processes, the number of multiplexed interferometric fiber hydrophone arrays can reach 100.

The scale of a TDM array is very limited, and the introduction of distributed amplification is an effective expansion method. During the propagation of the probe pulse along a fiber, distributed amplification can continuously or periodically transfer the energy of the pump light to the probe light, so that the probe light can travel a longer distance without exceeding the threshold of the non-linear effect. Figure 13 shows a system using distributed Raman amplification. In the system, the pump light pulse and probe light pulse are injected into a fiber hydrophone array concurrently. When the pump light and probe light meet, the stimulated Raman effect occurs, achieving distributed amplification [62]. Some researchers also propose that distributed erbium-doped amplifiers be integrated on the basis of distributed Raman amplifiers to further improve signal energy and increase the multiplexing scale to 64 TDM arrays.

On the basis of TDM, hybrid WDM and space division multiplexing can further expand the scale of fiber hydrophone arrays. As shown in Figure 14, dense WDM (DWDM) of 32 wavelengths can be implemented in the 1550 nm band. However, considering the cost of multiple lasers and the processing and layout of the sensor array, the multiplexing efficiency is not high [63]. In the SDM method, an optical switch is used to switch the connectivity state between a detection system and multiple hydrophones. Because the light source and detector are shared, the cost is significantly reduced, but the response frequency of the sensor is lower [64].

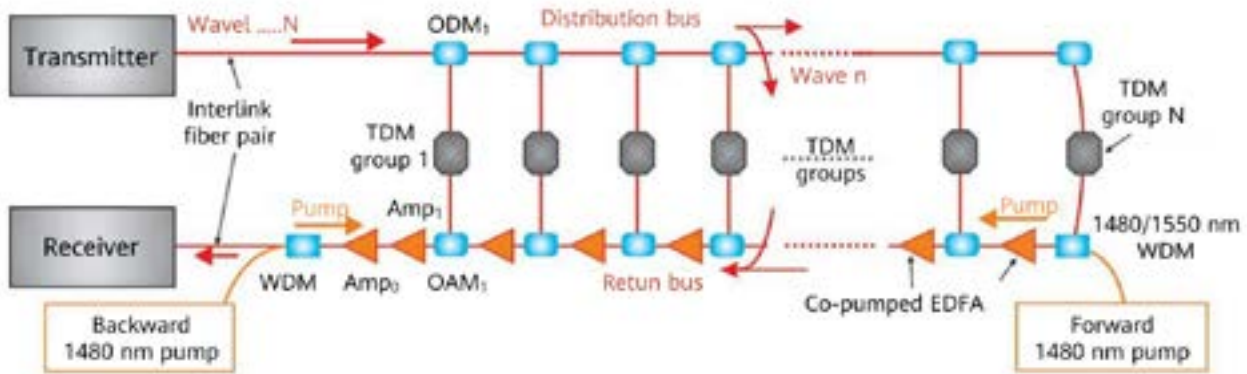


Figure 13 Interferometric fiber hydrophone array based on distributed amplification [62]

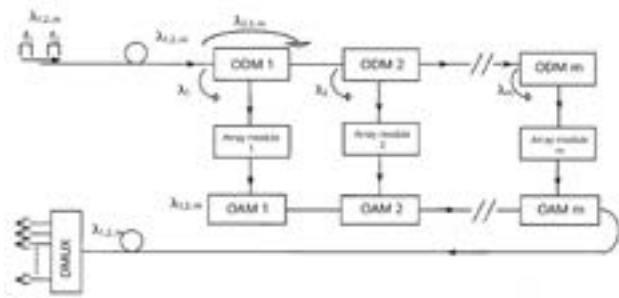


Figure 14 Fiber hydrophone array based on WDM [63]

Limited by the scale of system multiplexing, the research of interferometric fiber hydrophone arrays has not advanced significantly in the last ten years. In order to develop this technology, new lasers, detectors, signal processing technologies, and sensing mechanisms such as distributed acoustic wave detection are required.

4 Fully Distributed Fiber Sensor and Its Applications

The distributed fiber sensing technology is an important part of the fiber sensing field. Fibers implement both transmission and sensing, and collect information from millions of measurement points on a sensing fiber link, thereby achieving large-scale and long-distance sensing and monitoring. At present, distributed fiber sensing has been widely used in many fields such as perimeter security protection, pipeline safety, resource exploration, structural health monitoring, and medical testing, showing excellent engineering prospects.

Distributed fiber sensing technologies can be divided into interferometric distributed fiber sensing and scattering distributed fiber sensing. Interferometric distributed fiber sensing realizes synchronous sensing and positioning by constructing a composite interferometer. This technology has the advantages of high sensitivity, wide frequency response, and low cost, but it is susceptible to interference. In addition, the risk of positioning errors is large, and the structure of fiber is complex, so not much research is being conducted at present. In recent years, the mainstream distributed sensing technology has been the scattering distributed fiber sensing technology, that is, the backscattering light generated due to uneven media in the fiber, including the elastic Rayleigh scattering with unchanged frequency and the inelastic Brillouin scattering and Raman scattering with changed frequency, as shown in Figure 15. This technology uses optical time

domain reflectometry (OTDR), optical frequency domain reflectometry (OFDR), and optical correlation domain reflectometry (OCDR) to implement sensing and positioning. Generally, Rayleigh scattering distributed fiber sensors have great advantages in dynamic response to vibration and sound waves. Brillouin scattering distributed fiber sensors are usually used for the synchronous measurement of slowly changing signals such as temperature and strain. Raman scattering distributed fiber sensors are mainly used for distributed temperature measurement. The following describes backscattering distributed sensing technologies and typical applications of different sensing mechanisms.

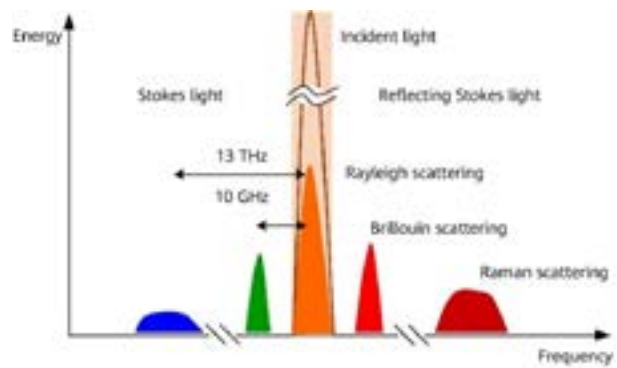


Figure 15 Scattering mechanism in quartz fibers

4.1 Rayleigh Scattering-based Distributed Fiber Sensing Technology and Its Applications in Acoustic Wave/Strain Measurement

Among fiber backscattering optical signals, Rayleigh scattering optical signals can account for more than 97% of the energy [65]. Compared with the other two scattering effects, Rayleigh scattering has a higher sensing SNR. The distributed sensing technology based on Rayleigh scattering implements distributed vibration/acoustic wave and strain sensing by measuring the polarization, intensity, and phase of backscattering optical signals.

4.1.1 Distributed Acoustic Sensing (DAS) Technology and Applications Based on Phase-Sensitive OTDR (ϕ -OTDR)

The distributed sensing technology based on phase detection has very high detection sensitivity. Currently, distributed sensing technologies based on Rayleigh optical phases can be classified into two types. One is the direct

detection of the backscattering light intensity, which can quickly respond to vibration events on fibers. However, because the response to external environment parameters is not linear, this technology is mainly used for distributed vibration sensing (DVS). The other is the demodulation of the phase information in a linear relationship with a measured parameter with high fidelity by constructing interferometric structures, which is mainly used for distributed acoustic sensing (DAS). In recent years, fiber DAS technology has gradually become a research hotspot for distributed sensing, and has developed fastest in the engineering application field.

4.1.1.1 Sensing Principles of Fiber DAS

Where DAS realizes acoustic wave sensing by measuring the optical phase change caused by the change in the axial strain of a fiber [66]. According to the photoelastic effect, the linear relationship between the axial strain of a fiber and the optical phase change is as follows:

$$\Delta\varphi = \beta \left[1 - \frac{n^2}{2}(P_{12} + 2P_{11}) \right] \Delta L \tag{5}$$

Where, β is the propagation constant of light, n is a refractive index of the fiber, P_{11} and P_{12} are the tensor coefficients of the fiber, and ΔL is the length change of the fiber. As shown in Figure 16, as long as the phase difference between points A and B of the probe light on the fiber can be extracted, the axial strain change of the fiber can be obtained. In this way, the quantitative sensing of the acoustic wave is achieved. In this process, high-fidelity fiber phase demodulation is the key to implementing high-performance DAS. Common optical phase demodulation technologies include the phase demodulation solution based on 3 x 3 couplers [67], phase generation carrier (PGC) solution [68], heterodyne/homodyne coherent detection solution [69], and phase demodulation solution based on linear sweep pulses [70]. The heterodyne coherent detection technology has better performance in SNR and noise suppression.

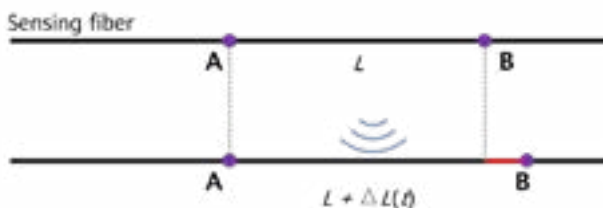


Figure 16 Principle of acoustic detection based on optical phase

4.1.1.2 NR Optimization for Scattering-Enhanced Fibers

Due to the randomness of Rayleigh scattering intensity and position in common single-mode fibers, DAS suffers from coherent fading and poor signal time consistency. To solve this problem, researchers use fibers with enhanced discrete scattering [71]. As shown in Figure 24, fibers of this type have a series of equally spaced scattering enhancement points (SEP) distributed axially along each fiber core. These SEPs have much stronger backscattering than their equivalent Rayleigh scattering points on a common fiber. Since they are also the sole determiners of the Rayleigh scattering signals, they overcome the randomness of the Rayleigh scattering intensity and position of an equivalent single-mode fiber. Keeping the spatial width of detection pulses smaller than the interval between adjacent SEPs eliminates interference fading caused by pulse overlapping, and allows for acoustic wave information with high stability and SNR.

Currently, we enhance the discrete scattering of a fiber using either of two methods. One is to introduce a periodic ultra-weak fiber Bragg grating (UWFBG) array [71, 72] into a fiber. To do this, ultraviolet light [1, 2, 73] or femtosecond laser [74–76] causes a permanent, periodic change in the fiber's refractive index, which in turn causes backward reflection of a specific wavelength bandwidth. However, in specific environments such as high temperature, low temperature, and high voltage, the UWFBG reflection spectrum may not match the detection wavelength, resulting in a sensing dead zone.

The other method of enhancing the discrete scattering is to introduce a local colorless weak reflection array [77–79]. Here, the fiber's local refractive index changes when it is exposed to ultraviolet light to form a colorless weak reflection point, which has stronger environmental adaptability.

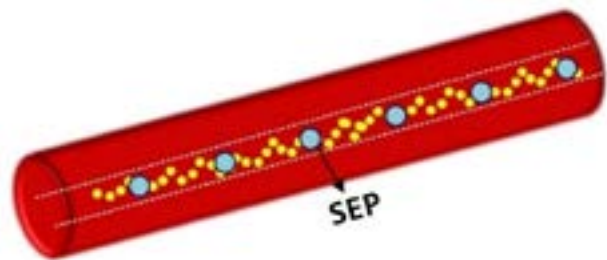


Figure 17 Structure of a fiber with enhanced discrete scattering [71]

By taking advantage of the high SNR and discretization of fibers with enhanced discrete scattering, researchers have applied these fibers to improving DAS performance, including in low-frequency phase drift compensation, polarization fading suppression, pulse width compression, and sampling frequency extension [80–87]. This key research aims to improve typical DAS performance from multiple dimensions, such as system noise floor, spatial resolution, and response frequency band.

4.1.1.3 Typical Applications of Fiber DAS

With the continuous improvement and development of the DAS system, researchers have explored its application in many engineering fields, including geological and resource exploration, structural health monitoring, and underwater detection.

Vertical seismic profiling (VSP) is a common seismic observation method for oil exploration. Traditional VSP uses point electronic detectors that are vulnerable to high temperature and pressure, making it inefficient for large-scale measurement. DAS technology makes fiber logging a possible alternative. A microstructure fiber with enhanced scattering has stable backscattering optical sensing signals with a high SNR, which increases the stability of phase demodulation, fidelity of seismic waves, and consistency between signals from multiple channels. The research

team of Huazhong University of Science and Technology performed a walkaway VSP experiment [88] using a self-fabricated microstructure fiber DAS system. VSP images at "zero offset" and "non-zero offset" positions were recorded, in which the direct and reflected waves with high SNR were observed.

The health of infrastructure such as railways, tunnels, and pipelines is vital to personal and economic safety. Fiber DAS systems have been widely used in structural health monitoring. Zhaoqiang Peng et al. tracked how acoustic waves propagate in pipelines using a DAS system with high SNR, and analyzed the DAS data using neural network machine learning to monitor and identify external intrusion events and internal corrosion status of the pipelines, as shown in Figure 18 [89]. In 2019, Qizhen Sun et al. detected railway defects using fiber DAS [90]. They deployed sensing fibers at the waist of the tracks to record signals from the acoustic emission generated and propagated when train wheels pass a defective point, and used a positioning algorithm to accurately locate the defect. The team then applied this system to the intelligent monitoring of the steel ring structures that reinforce tunnels [91]. Likewise, they deployed sensing fibers on the steel ring structures to track the resonance of an active sound source. When the gap between the reinforced steel ring and the tunnel body fails, a cavity will change the resonant wave characteristics. Such changes can be analyzed by back propagation (BP) neural network machine learning to accurately identify the severity of the cavity.

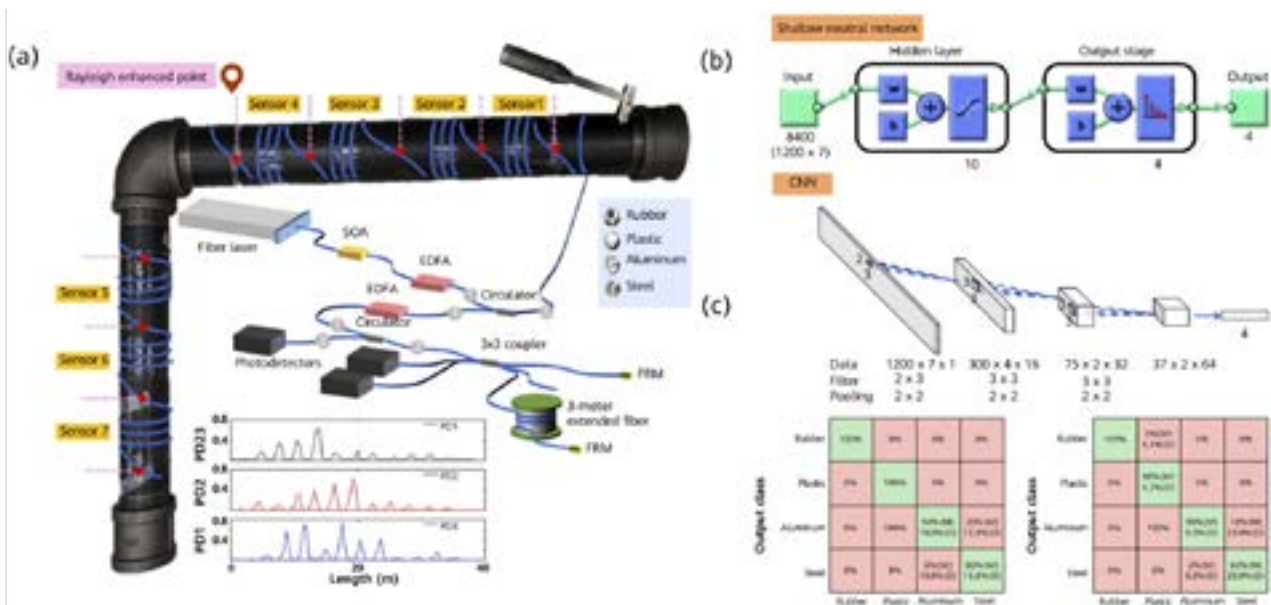


Figure 18 Pipe measurement and results [89]: (a) DAS system structure and layout of fibers with enhanced scattering; (b) Schematic diagram of the neural network; (c) Confusion matrix of four types of events

In recent years, fiber DAS has been introduced into the field of underwater detection to construct a new generation of distributed fiber hydrophones. Because fibers are insensitive to lateral pressure, traditional straight-through optical cables are only sensitive to a sound pressure up to -210 dB *re* rad/ μ Pa. To improve hydrophone sensitivity, the research team of Huazhong University of Science and Technology designed a lightweight fully distributed hydrophone cable [92] using two sensitivity enhancing mechanisms: internal elastic core layer and fiber spiral winding. As shown in Figure 19, the sensitivity now reaches -137 dB *re* rad/ μ Pa, with flat frequency response range of 5–2000 Hz. Lake testing shows that this improved hydrophone can clearly track the movement of ships and the radiation noise of small submarines.

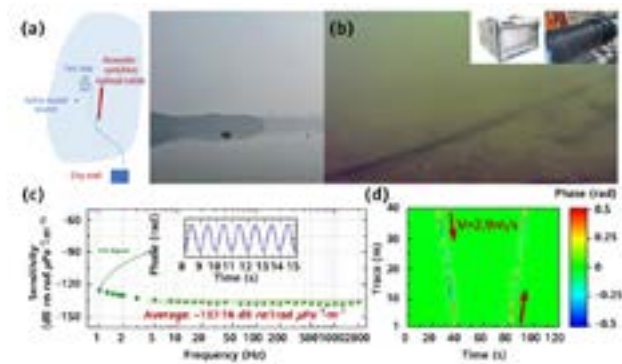


Figure 19 Distributed hydrophone optical cable: (a) Lake test; (b) Underwater layout diagram; (c) Sound pressure frequency response curve; (d) Recorded vessel movement signal

4.1.2 Principle and Applications of OFDR Distributed Shape Sensing

OFDR shares the same basic principle of frequency-modulated continuous wave (FMCW) technology used in the microwave field. A laser with linearly tunable wavelengths is the usual light source. The continuous light emitted by the laser is divided into two beams: one reference light, and one probe light emitted to the fiber under test (FUT). Rayleigh backscattering signals from FUT interfere with the reference

light after linear frequency sweeping, generating a beat frequency signal that is proportional to the time difference. The strain of each position on the fiber is obtained from the spectral offset. To map the position of Rayleigh scattering on the fiber and the beat frequency, we use Fourier transform of the beat frequency signals from the detector. The spatial resolution of an OFDR system depends on the frequency sweep range of the light source and spectrum analysis precision, but not by detector bandwidth. Therefore, an OFDR system has a high spatial resolution precise to the micrometer.

Using fiber OFDR, three-dimensional (3D) shape sensing is realized. Figure 20 shows the working principle and process. First, the OFDR system demodulates the strain data from the Rayleigh scattering spectra of multiple fibers. It then measures the fiber shape based on the different strain responses of deformed fiber cores, performs model mapping for the different strain changes to obtain shape parameters such as bending curvature and bending angle, and then uses a reconstruction algorithm to obtain the 3D shape.

Shape sensing using fiber OFDR has obvious application advantages in medical, energy, defense, aerospace, and structural safety monitoring. For example, navigation is key in minimally invasive interventional surgery, and needs to provide a resolution in millimeters when measuring the shape of the apparatus used [93]. In 2010, the company LUNA applied to patent the use of spiral multi-core fibers as shape sensors. The American Intuitive Surgical Company then applied this technology to Ion which is a minimally invasive intervention navigation system. As shown in Figure 21, fiber-optic shape sensing allows for active control of the catheter position throughout the navigation and biopsy process. At the target nodule, the catheter locks in place to provide the stability required for the precise placement of the biopsy tool [94]. In 2017, the Polytechnic Institute of Montreal Canada used scattering-enhanced fibers for higher SNR, improving the average measurement accuracy of a puncture needle by 47% [95]. In addition, 3D shape sensing of catheters were realized in an *in vitro* vascular model and in live pigs [96]. In 2015, NASA Armstrong Flight Research

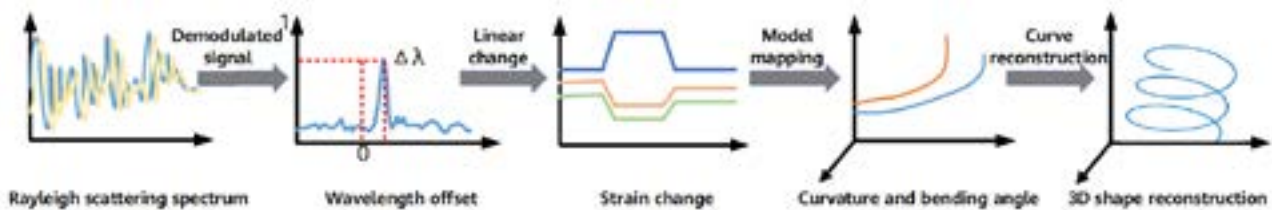


Figure 20 Working principle of fiber OFDR-based 3D shape sensing

Center furthered the application of fiber 3D shape sensing in the deformation measurement of rocket delivery systems. In addition, the American Intuitive Surgical Company [97] and the Korean Academy of Sciences and Technology [98] used fiber shape sensing in smart wearables for pose monitoring, such as finger motion capture.

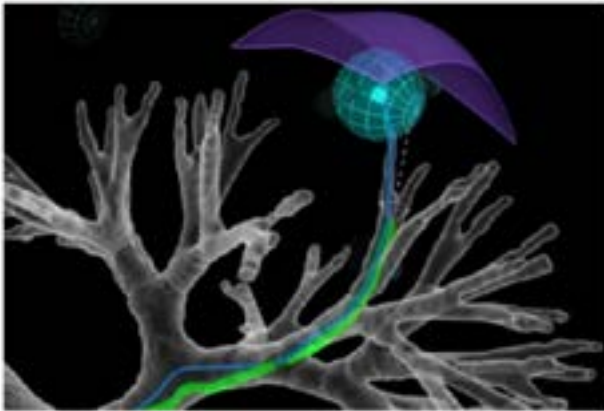


Figure 21 Sampling of lung tissue by biopsy with an Ion system integrated with a fiber-optic shape sensor [94]

4.2 Distributed Optical Fiber Sensing Based on Brillouin Scattering and its Applications

Brillouin scattering is a fiber nonlinear effect caused by frequency shifts from inelastic collisions between incident photons and acoustic phonons. Brillouin scattering light is a series of frequencies distributed by the Lorentz spectral line. The frequency difference between the center of the spectral line and the incident light is the Brillouin frequency shift (BFS), which is linearly sensitive to both temperature and strain.

4.2.1 Distributed Optical Fiber Sensing Based on Brillouin Backscattering

The two technology types are fiber Brillouin optical time domain reflectometer (BOTDR) based on spontaneous Brillouin scattering, and fiber Brillouin optical time domain analyzer (BOTDA) based on stimulated Brillouin scattering. BOTDR is a distributed sensing system where common coherence detection is either autocorrelation homodyne detection or microwave heterodyne detection. BOTDR features single-ended pumping and simple structure, but

has a low SNR. BOTDA is a distributed sensing system where scanning the Brillouin gain spectrum (BGS) with frequency distribution provides the BFS indirectly and demodulates disturbance information.

The most important accuracy determiner of both BOTDR and BOTDA (BOTDR/A) sensing is the accurate extraction of BFS from the BGS curve. Traditionally, it is extracted using Lorentz Curve Fitting (LCF), but this solution requires an initial LCF value. Therefore, the BFS extraction accuracy deteriorates rapidly when the SNR is low. In addition, fiber doping and the cross-sectional structure cause the BGS in practice to deviate from strict Lorentz distribution, resulting in fitting errors. To solve this problem, many BFS extraction technologies propose to improve the sensing accuracy of BOTDR/A by:

- Boosting injection power. In 2018, Nageswara Lalam et al. used a multi-wavelength method to improve BOTDR, with a three-wavelength pump to increase the SNR by 4.85 dB [99].
- Coded pulse pumping. In 1993, M.D. Jones improved row performance using Simplex encoding [100]. As shown in Figure 22, in 2011, M. Taki et al. combined differential pulse pairs and Simplex coding in a BOTDA system, achieving 1.1 MHz BFS extraction precision [101].
- Similarity matching algorithm. The BGS without sensing signals is used as a reference spectrum to calculate similarity between the reference and sensing BGS, eliminating noise, improving SNR, and achieving 0.92 MHz BFS extraction precision [102].
- Machine learning. Thanks to recent AI developments, various methods including convolutional neural network (CNN) and support vector machine (SVM) are used for BFS extraction. CNN-based extraction not only improves precision, but also outperforms the traditional Lorentz curve fitting in processing efficiency [103].

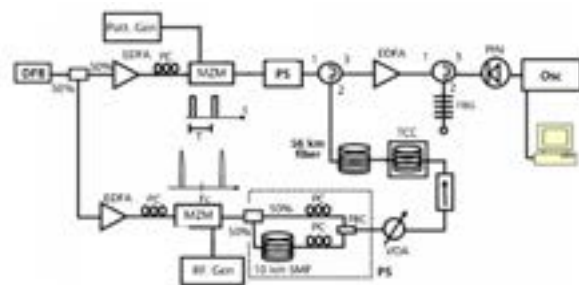


Figure 22 Block diagram of a BOTDA system based on differential pulse pairs and Simplex coding [101]

Spatial resolution is also important for BOTDR/A. Although decreasing the pulse width can reduce the spatial resolution, it also reduces the pulse energy. A pulse width smaller than the phonon lifetime triggers BGS broadening, which seriously affects detection precision. There are two ways to improve the spatial resolution:

- Differential pulse pairs. The resolution improves when the pulse width is not less than the phonon lifetime. Studies have reported a resolution as low as 41 cm within the 56 km sensing range [101].
- Signal processing algorithm. In 2013, a Nanjing University team proposed an iterative subdivision method using Brillouin spectrum analysis [104]. The original spatial resolution δz is subdivided m times to obtain m sub-segments. As shown in Figure 23, the Brillouin spectrum at the original measurement location z_0 is considered to be the superposition of the m sub-spectrums, and recursive iteration improves resolution by m times.

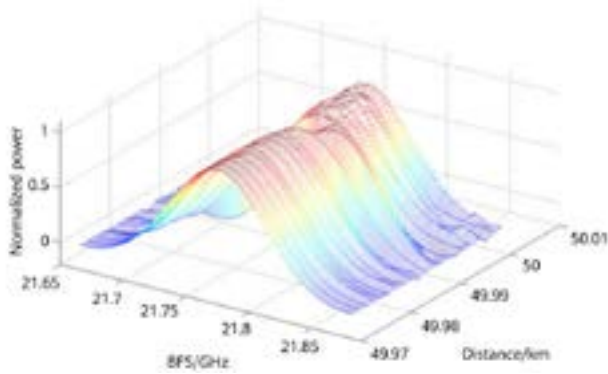


Figure 23 Brillouin spectrum after iterative subdivision

As described above, BOTDR/A measurement is more precise but slower. It is thus better for long-distance temperature/strain sensing that needs relative precision but not strict real-time performance. Moffat from the University of Chile attached a BOTDR sensing fiber to a PVC pipe in the lab, simulated displacement monitoring for a mine wall, and then monitored field strain in a mine tunnel [105]. Cheng from Nanjing University buried BOTDR sensing fibers in the rock and soil of a coal mine to monitor the overburden deformation in real time. By collecting the data of strain distribution that varies during the mining process, the team created an overburden deformation pattern and used the data in mine safety pre-evaluation and accident prevention [106].

4.2.2 Distributed Optical Fiber Sensing Based on Forward Brillouin Scattering

Forward stimulated Brillouin scattering (FSBS) is a phenomenon of energy transfer. This transfer is between codirectional pump/Stokes light and transverse acoustic waves and is caused by electrostriction and elasto-optic effects. In 2016, the Zadok A team first proposed using FSBS to measure liquid surrounding a single-mode fiber and using the R0,m radiation acoustic waves for sensing [107]. The wave duration reflects the difference between the fiber's external material and its acoustic wave impedance – the larger the difference, the higher the wave reflectivity at the boundary, and the longer the service life.

In 2018, the FSBS-based distributed sensing concept [108-109] attracted wide attention and pilot application in measuring fiber diameter [110], sound velocity [111], stress [112], and temperature [113]. The Zadok A and Thvenaz L teams independently reported on an FSBS-based distributed liquid sensor. Figure 24 shows the sensing principle diagram [108] proposed by Zadok A. A dual-frequency long pulse is injected into one end of a fiber and the distributed FSBS gain spectrum is obtained by scanning the FSBS frequency and Rayleigh backscattering power. Subsequently, researchers perform phase encoding on the pulse, wavelength tuning on the laser, and distributed measurement of air, alcohol, and water by averaging out coherent Rayleigh scattering noise. The Thvenaz L team used a dual-frequency long pulse to excite a sound wave field, used short read pulses to detect accumulated phase modulation, and used differential demodulation to obtain a distributed FSBS spectrum [109]. FSBS is a new methodology for distributed fiber sensing, but due to the limitations of transverse acoustic duration and

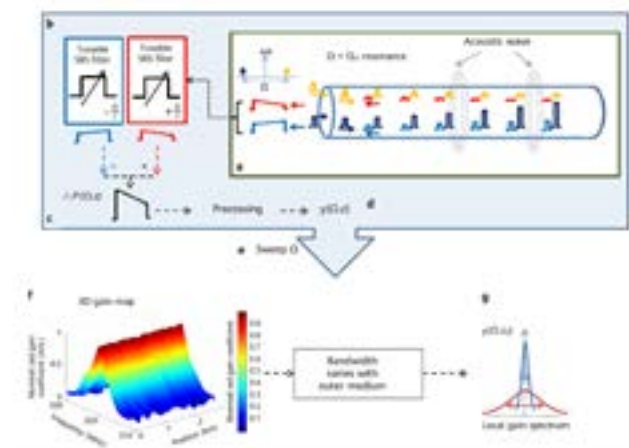


Figure 24 FSBS liquid sensing principle [108]

SNR, resolution is only precise to tens of meters, and system stability and accuracy need to be improved.

4.3 Distributed Optical Fiber Sensing Based on Raman Scattering and its Applications in Temperature Measurement

Raman scattering is an inelastic collision between incident photons and optical phonons in a fiber, which generates both Stokes light and anti-Stokes light of a lower and higher frequency than the incident light respectively [114]. The intensity of anti-Stokes light is sensitive to temperature, while Stokes light is a relatively stable reference, meaning that Raman scattering distributed fiber temperature measurement using OTDR is possible. This system is mature and has been fully commercialized into many products widely used in temperature anomaly detection for gas pipelines and power cables.

5 Summary and Prospects

This paper mainly comprises three parts: an overview of point, quasi-distributed, and fully distributed fiber sensors; an elaboration on the progress on single-point and array sensors based on fiber grating, Fabry-Perot Interferometric, and spectral absorption; and finally a description of the basic principles, key technologies, system research, and application progress of distributed scattering fiber sensors. Up to now, optical fiber sensing technology in China is in a high-speed development period, and can still be improved from several aspects. First, for different types of fiber sensors, sensing performance for sensitivity, noise level, and measurement bandwidth would be better with new optical sensing mechanisms developed on existing sensing principles. In addition, general big data processing algorithms for multi-field applications can be developed with machine learning. Second, special sensing fibers and optical cables are required for specific scenarios. Optical integrated modules and devices that work with such sensing fibers need to be developed for fiber sensing systems that are economical, reliable, and integrated. Finally, optical fiber sensing must be aligned with engineering application and market requirements in various fields such as medical biology, resource exploration, and military defense. Optical

fiber sensing is a promising field with unique advantages over traditional electromagnetic sensors, with the potential to catalyze a technological revolution.

References

- [1] Hill K O, Fujii Y, Johnson D C, et al. Photosensitivity in optical fiber waveguides: Application to reflection filter fabrication. *Applied Physics Letters*, 1978, 32(10): 647-649.
- [2] Melts G, Mory W W, Glenn W H. Formation of Bragg gratings in optical fibers by transverse holographic method. *Optics Letters*, 1989, 14(15): 823-825.
- [3] Hill K O. Bragg gratings fabricated in monomode Photosensitivity optical fiber by UV exposure through a Phase mask. *Applied Physics Letters*, 1993, 62(10): 1035-1037.
- [4] Majumder M, Gangopadhyay T K, Chakraborty A K. Fibre Bragg gratings in structural health monitoring—Present status and applications. *Sensors & Actuators A Physical*, 2008, 147(1): 150-164.
- [5] Shu X, Lin Z, Bennion I. Sensitivity characteristics of long-period fiber gratings. *Journal of Lightwave Technology*, 2002, 20(2): 255-266.
- [6] James S W, Tatam R P. Optical fibre long-period grating sensors: Characteristics and application. *Measurement Science and Technology*, 2003, 14(5).
- [7] Janczuk-Richter M, Dominik M, Ro Niecka E. Long-period fiber grating sensor for detection of viruses. *Sensors & Actuators B Chemical*, 2017, 250:32-38.
- [8] W. Gan, Z. Xu, Y. Li, W. Bi, L. Chu, Q. Qi, Y. Yang, P. Zhang, N. Gan, S. Dai, and T. Xu. Rapid and sensitive detection of *Staphylococcus aureus* by using a long-period fiber grating immunosensor coated with egg yolk antibody. *Biosensors and Bioelectronics*, 2022, 199(1): 113860.
- [9] Tuan Guo, Fu Liu, Liyang Shao. Tilted fiber grating sensor. *Journal of Applied Science*, 2018, 36(1): 29.

- [10] T. Guo, F. Liu, Y. Liu, N.-K. Chen, B.-O. Guan, and J. Albert. In-situ detection of density alteration in non-physiological cells with polarimetric tilted fiber grating sensors. *Biosensors and Bioelectronics*, 2014, 55(9): 452-458.
- [11] Zhang Z, Guo T, Zhang X, et al. Plasmonic fiber-optic vector magnetometer. *Applied Physics Letters*, 2016, 108(10):101105.
- [12] Huang Z Y, Zhu Y Z, Chen X P, Wang A B. Intrinsic Fabry-Perot fiber sensor for temperature and strain measurements. *Ieee Photonics Technology Letters*, 2005, 17(11): 2403-2405.
- [13] Zhang L, Jiang Y, Gao H, Jia J, Cui Y, Wang S, et al. Simultaneous Measurements of Temperature and Pressure With a Dual-Cavity Fabry-Perot Sensor. *IEEE Photonics Technology Letters*, 2019, 31(1): 106-109.
- [14] Xu J C, Wang X W, Cooper K L, Wang A B. Miniature all-silica fiber optic pressure and acoustic sensors. *Optics Letters*, 2005, 30(24): 3269-3271.
- [15] Yu Q, Zhou X. Pressure sensor based on the fiber-optic extrinsic Fabry-Perot interferometer. *Photonic Sensors*, 2011, 1(1): 72-83.
- [16] Zhixia Zheng, Yuanqing Huang, Yongjian Feng. Optical fiber MEMS pressure sensor for distributed measurement of denture pressure. *Photoelectron laser*, 2012, 23(12): 2261-2266.
- [17] Yang Y, Saurabh S, Ward J M, Chormaic S N. High-Q, ultrathin-walled microbubble resonator for aerostatic pressure sensing. *Optics Express*, 2016, 24(1): 294-299.
- [18] Wu X, Jan C, Solgaard O. Single-Crystal Silicon Photonic-Crystal Fiber-Tip Pressure Sensors. *Journal of Microelectromechanical Systems*, 2015, 24(4): 968-975.
- [19] Inaudi D, Pinet E. Large-volume Fabry-Pérot fiber-optic sensors production for medical devices and industrial applications. in: 26th International Conference on Optical Fiber Sensors, Lausanne, 24 Sep. 2018, Proceedings of the Optical Society of America, 2018:TuE32.
- [20] Wang A B. Optical fiber sensors for energy-production and energy intensive industries. in: Conference on Advanced Sensor Systems and Applications, Shanghai, China, 15-18 Oct. 2002, Proceedings of the SPIE, 2002:377-381.
- [21] Zhang E Z, Beard P C. Characteristics of optimized fibre-optic ultrasound receivers for minimally invasive photoacoustic detection. in: Conference on Photons Plus Ultrasound - Imaging and Sensing, San Francisco, CA, 08-10 Feb. 2015, Proceedings of the SPIE, 2015:932311.
- [22] Zhang W, Chen F, Ma W, Rong Q, Qiao X, Wang R. Ultrasonic imaging of seismic physical models using a fringe visibility enhanced fiber-optic Fabry-Perot interferometric sensor. *Optics Express*, 2018, 26(8): 11025-11033.
- [23] Chen B, Chen Y, Ma C. Photothermally tunable Fabry-Perot fiber interferometer for photoacoustic mesoscopy. *Biomedical Optics Express*, 2020, 11(5): 2607-2618.
- [24] Fischer B. Optical microphone hears ultrasound. *Nature Photonics*, 2016, 10(6): 356-358.
- [25] Alcoz J J, Lee C E, Taylor H F. Embedded fiber-optic Fabry-Perot ultrasound sensor. *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, 1990, 37(4): 302-306.
- [26] Morris P, Hurrell A, Shaw A, Zhang E, Beard P. A Fabry-Perot fiber-optic ultrasonic hydrophone for the simultaneous measurement of temperature and acoustic pressure. *Journal of the Acoustical Society of America*, 2009, 125(6): 3611-3622.
- [27] Guggenheim J A, Li J, Allen T J, Colchester R J, Noimark S, Ogunlade O, et al. Ultrasensitive plano-concave optical microresonators for ultrasound sensing. *Nature Photonics*, 2017, 11(11): 714-719.
- [28] Preisser S, Rohringer W, Liu M, Kollmann C, Zotter S, Fischer B, et al. All-optical highly sensitive akinetic sensor for ultrasound detection and photoacoustic imaging. *Biomedical Optics Express*, 2016, 7(10): 4171-4186.
- [29] H. Inaba, T. Kobayasi, M. Hirama, and M. Hamza. Optical-fibre network system for air-pollution

- monitoring over a wide area by optical absorption method. *Electronics Letters*, 1979, 15: 749-751.
- [30] Li Bin, Zheng Chuantao, Liu Huifang, et al. Development and measurement of a near-infrared CH₄ detection system using 1.654 μm wavelength-modulated diode laser and open reflective gas sensing probe[J]. *Sensors and Actuators B: Chemical*, 2016, 225: 188-198
- [31] J. Reid and D. Labrie. Second-harmonic Detection with Tunable Diode Lasers – Comparison of Experiment and Theory. *Applied Physics B Photophysics and Laser Chemistry*, 1981, 26(3): 203-210.
- [32] Li J, Parchatka U, Fischer H. Development of field-deployable QCL sensor for simultaneous detection of ambient N₂O and CO. *Sensors & Actuators B Chemical*, 2013, 182:659-667.
- [33] U. Willer, D. Scheel, I. Kostjucenko, C. Bohling, W. Schade, and E. Faber. Fiber-optic evanescent-field laser sensor for in-situ gas diagnostics. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 2002, 58(11): 2427-2432.
- [34] Chengju Ma, Weifeng Xu, Jiamei Li, Mei Yang, Keyang Liu, Mengting Li. Research progress of micro nanofiber evanescent field effect gas sensor. *Progress in Laser and Optoelectronics*, 2015, 52 (10): 26-31.
- [35] F. Benabid, F. Couny, J. C. Knight, T. A. Birks, and P. S. J. Russell. Compact, stable and efficient all-fibre gas cells using hollow-core photonic crystal fibres. *Nature*, 2005, 434: 488-491.
- [36] S. Olyaei, A. Naraghi, and V. Ahmadi. High sensitivity evanescent-field gas sensor based on modified photonic crystal fiber for gas condensate and air pollution monitoring. *Optik*, 2014, 125(1): 596-600.
- [37] Chen N K, Hsieh Y H, Lee Y K. Tapered fiber Mach-Zehnder interferometers for vibration and elasticity sensing applications[J]. *Optics express*, 2013, 21(9): 11209-11214.
- [38] Li J, Chen J, Xu F. Sensitive and wearable optical microfiber sensor for human health monitoring[J]. *Advanced Materials Technologies*, 2018, 3(12): 1800296.
- [39] Zhu H T, Zhan L W, Dai Q, et al. Self-Assembled Wavy Optical Microfiber for Stretchable Wearable Sensor[J]. *Advanced Optical Materials*, 2021, 9: 2002206.
- [40] Li L, Song C, Liu Y, et al. Wearable and Alignment-free Optical Microfiber Device for Human Health Monitoring[C]//2021 19th International Conference on Optical Communications and Networks (ICOON). IEEE, 2021: 1-3.
- [41] Zhang L, Pan J, Zhang Z, et al. Ultrasensitive skin-like wearable optical sensors based on glass micro/nanofibers[J]. *Opto-Electronic Advances*, 2020, 3(3): 190022-1-190022-7.
- [42] Pan J, Zhang Z, Jiang C, et al. A multifunctional skin-like wearable optical sensor based on an optical micro-/nanofibre[J]. *Nanoscale*, 2020, 12(33): 17538-17544.
- [43] Li Y, Tan S, Yang L, et al. Optical Microfiber Neuron for Finger Motion Perception[J]. *Advanced Fiber Materials*, 2021: 1-9.
- [44] Sun D, Guo T, Ran Y, et al. In-situ DNA hybridization detection with a reflective microfiber grating biosensor[J]. *Biosensors and Bioelectronics*, 2014, 61: 541-546.
- [45] Zhang L, Wang P, Xiao Y, et al. Ultra-sensitive microfiber absorption detection in a microfluidic chip[J]. *Lab on a Chip*, 2011, 11(21): 3720-3724.
- [46] Liu G, Li K. Micro/nano optical fibers for label-free detection of abrin with high sensitivity[J]. *Sensors and Actuators B: Chemical*, 2015, 215: 146-151.
- [47] Tian Y, Wang W, Wu N, et al. Tapered optical fiber sensor for label-free detection of biomolecules[J]. *Sensors*, 2011, 11(4): 3780-3790.
- [48] Silva S, Coelho L, Almeida J M, et al. H₂ sensing based on a Pd-coated tapered-FBG fabricated by DUV femtosecond laser technique[J]. *IEEE photonics technology letters*, 2013, 25(4): 401-403.
- [49] Fu H, Jiang Y, Ding J, et al. Zinc oxide nanoparticle incorporated graphene oxide as sensing coating for interferometric optical microfiber for ammonia gas detection[J]. *Sensors and Actuators B: Chemical*, 2018, 254: 239-247.

- [50] Zibaii M I, Taghipour Z, Saeedian Z, et al. Kinetic study for the hybridization of 25-mer DNA by nonadiabatic tapered optical fiber sensor[C]//2011 Asia Communications and Photonics Conference and Exhibition (ACP). IEEE, 2011: 1-6.
- [51] Huang Y, Yu B, Guo T, et al. Ultrasensitive and in situ DNA detection in various pH environments based on a microfiber with a graphene oxide linking layer[J]. RSC advances, 2017, 7(22): 13177-13183.
- [52] Li K, Liu G, Wu Y, et al. Gold nanoparticle amplified optical microfiber evanescent wave absorption biosensor for cancer biomarker detection in serum[J]. Talanta, 2014, 120: 419-424.
- [53] Yuyuan Hu. Research on high-speed demodulation method and key technology of large-capacity fiber grating sensing network[D]. Wuhan: Wuhan University of Technology.
- [54] Wang W, Gong J, Dong B, et al. A Large Serial Time-Division Multiplexed Fiber Bragg Grating Sensor Network[J]. Lightwave Technology Journal of, 2012, 30(17):p.2751-2756.
- [55] Lindner E, Canning J, Chojetzki C, et al. Post-hydrogen-loaded draw tower fiber Bragg gratings and their thermal regeneration[J]. Applied Optics, 2011, 50(17):2519-22.
- [56] Guo H, Tang J, Li X, et al. On-line writing identical and weak fiber Bragg grating arrays[J]. Chinese Optics Letters, 2013(3):4.
- [57] Deming Liu, Tao He, Zhijie Xu. New microstructure fiber distributed acoustic wave sensing technology and its applications[J]. Journal of Applied Science, 2020,38(02):296-309.
- [58] Baijie Xu, Jun He, Bin Du, et al. Femtosecond laser point-by-point inscription of an ultra-weak fiber Bragg grating array for distributed high-temperature sensing[J]. Optics Express, 2021 29:32615-32626.
- [59] Gopinath R, Arora P, Gandhi G, et al. Broadband fiber optic hydrophone sensors for sub-millimeter ultrasound resolutions. IEEE, 2008.
- [60] Bucaro, J, A. Fiber-optic hydrophone[J]. Journal of the Acoustical Society of America, 1977.
- [61] Cranch G A, Nash P J, Kirkendall C K. Large-scale remotely interrogated arrays of fiber-optic interferometric sensors for underwater acoustic applications[J]. IEEE Sensors Journal, 2003, 3(1): 19-30.
- [62] Liao Y, Austin E, Nash P J, et al. Highly Scalable Amplified Hybrid TDM/DWDM Array Architecture for Interferometric Fiber-Optic Sensor Systems[J]. J. Lightwave Technol., 2013, 31(6): 882-888.
- [63] Cranch G A, Nash P J. High multiplexing gain using TDM and WDM in interferometric sensor arrays[J]. Proceedings of SPIE the International Society for Optical Engineering, 1999, 3860.
- [64] Yongming Hu, Zhou Meng, Shuidong Xiong, Yang Liu, Xueliang Zhang, Guoqiang Fei, Zhaoxia Wang. Development of interferometric all polarization maintaining fiber hydrophone array[J]. Journal of Acoustics, 2003(02):155-158.
- [65] Bolognini G, Hartog A. Raman-based fibre sensors: Trends and applications. Optical Fiber Technology, 2013, 19(6): 678-688.
- [66] Juan, Juarez C, Eric, et al. Distributed Fiber-Optic Intrusion Sensor System[J]. Journal of Lightwave Technology, 2005, 23(6): 2081-2087.
- [67] Zhang X, Sun Z, Shan Y, et al. A high performance distributed optical fiber sensor based on Φ -OTDR for dynamic strain measurement[J]. IEEE Photonics Journal, 2017, 9(3): 1-12.
- [68] Fang G, Xu T, Feng S, et al. Phase-sensitive optical time domain reflectometer based on phase-generated carrier algorithm[J]. Journal of lightwave technology, 2015, 33(13): 2811-2816.
- [69] Pan Z, Liang K, Ye Q, et al. Phase-sensitive OTDR system based on digital coherent detection[C]// Asia Communications and Photonics Conference and Exhibition. Optical Society of America, 2011: 831105.
- [70] Chen D, Liu Q, Fan X, et al. Distributed fiber-optic acoustic sensor with enhanced response bandwidth and high signal-to-noise ratio[J]. Journal of Lightwave Technology, 2017, 35(10): 2037-2043.
- [71] Fan A, Sun Q, Wei Z, et al. Wideband Fully-Distributed Vibration Sensing by Using UWFBG

- Based Coherent OTDR[C]// 2017 Optical Fiber Communications Conference and Exhibition (OFC). IEEE, 2017: 1-3.
- [72] Huang J B, Ding P, Tang J S. Progress in Fabrication, Demodulation and Application of Weak-Reflection Fiber Bragg Grating Array[J]. *Laser & Optoelectronics Progress*, 2021, 58(17): 1700005.
- [73] Redding B, Murray M J, Donko A, et al. Low-noise distributed acoustic sensing using enhanced backscattering fiber with ultra-low-loss point reflectors[J]. *Optics express*, 2020, 28(10): 14638-14647.
- [74] Zheng Y, Yu H, Guo H, et al. Analysis of the Spectrum Distortions of Weak Fiber Bragg Gratings Fabricated In-Line on a Draw Tower by the Phase Mask Technique[J]. *Journal of Lightwave Technology*, 2015, 33(12): 2670-2673.
- [75] Martinez A, Dubov M, Khrushchev I, et al. Direct writing of fibre Bragg gratings by femtosecond laser[J]. *Electronics Letters*, 2004, 40(19): 1170-1172.
- [76] Ziyong Chen, Jun He, Xizhen Xu. Fabrication of fiber Bragg grating high temperature sensor array by femtosecond laser point-by-point method[J]. *Journal of Optics*, 2021, 41(13): 1306002.
- [77] Junbin Huang, Peng Ding, Jinsong Tang. Progress in fabrication, demodulation, and application of weak reflection fiber grating array[J]. *Progress in Laser and Optoelectronics*, 2021, 58(17): 1700005.
- [78] Wu M, Fan X, Zhang X, et al. Frequency Response Enhancement of Phase-sensitive OTDR for Interrogating Weak Reflector Array by Using OFDM and Vernier Effect[J]. *Journal of Lightwave Technology*, 2020, PP(99): 1-1.
- [79] Hicke K, Eisermann R, Chruscicki S. Enhanced Distributed Fiber Optic Vibration Sensing and Simultaneous Temperature Gradient Sensing Using Traditional C-OTDR and Structured Fiber with Scattering Dots[J]. *Sensors*, 2019, 19(19): 4114.
- [80] Wu M, Fan X, Liu Q, et al. Highly sensitive quasi-distributed fiber-optic acoustic sensing system by interrogating a weak reflector array[J]. *Optics letters*, 2018, 43(15): 3594-3597.
- [81] Liu T, Li H, Ai F, et al. Ultra-high Resolution Distributed Strain Sensing based on Phase-OTDR[C]//proceedings of the 2019 Optical Fiber Communications Conference and Exhibition (OFC). IEEE:1-3
- [82] Wang F, Liu Y, Wei T, et al. Polarization fading elimination for ultra-weak FBG array-based Phi-OTDR using a composite double probe pulse approach[J]. *Optics express*, 2019, 27(15): 20468-20478.
- [83] Liu T, Li H, He T, et al. Fading Noise Free Distributed Acoustic Sensor Assisted with Double Wavelength lasers[C]//proceedings of the CLEO: Science and Innovations. Optical Society of America:JW2E. 10.
- [84] Wu M, Fan X, Liu Q, et al. Quasi-distributed fiber-optic acoustic sensing system based on pulse compression technique and phase-noise compensation[J]. *Optics Letters*, 2019, 44(24): 5969-5972.
- [85] Yang W, Fu X, Wang J, et al. Demodulation Method of Identical Ultra-Weak Fiber Bragg Grating Array Based on Golay Code[C]// 2019 11th International Conference on Measuring Technology and Mechatronics Automation (ICMTMA). IEEE, 2019: 256-260.
- [86] Zhang Y X, Fu S Y, Chen Y S, et al. A visibility enhanced broadband phase-sensitive OTDR based on the UWFBG array and frequency-division-multiplexing[J]. *Optical Fiber Technology*, 2019, 53: 101995.
- [87] Li, H., Fan, C Z, Liu, T. et al. Time-slot multiplexing based bandwidth enhancement for fiber distributed acoustic sensing. *Science China Information Sciences*, 65, 119303 (2022).
- [88] Li Y P, Karrenbach M, Ajo J. Distributed Acoustic Sensing in Geophysics: Methods and Applications [M]. American Geophysical Union, 2022.
- [89] Peng Z, Jian J, Wen H, et al. Distributed fiber sensor and machine learning data analytics for pipeline protection against extrinsic intrusions and intrinsic corrosions[J]. *Optics Express*, 2020, 28(19): 27277-27292.

- [90] Fan C, Ai F, Liu Y, et al. Rail Crack Detection by Analyzing the Acoustic Transmission Process Based on Fiber Distributed Acoustic Sensor[C]// Optical Fiber Communication Conference. Optical Society of America, 2019: Th2A. 17.
- [91] Hu D, Tian B, Li H, et al. Intelligent Structure Monitoring for Tunnel Steel Loop Based on Distributed Acoustic Sensing[C]//proceedings of the CLEO: Applications and Technology. Optical Society of America:ATh1S. 4.
- [92] Chen J, Li H, Liu T, et al. Fully Distributed Hydroacoustic Sensing based on Lightweight Optical Cable assisted with Scattering Enhanced Fiber[C]//proceedings of the 2021 Optical Fiber Communications Conference and Exhibition (OFC). IEEE:1-3.
- [93] Waltermann C, Koch J, ANGELMAHR M, et al. Fiber optical 3D Shape Sensing[M]// Planar Waveguides and other Confined Geometries. New York, NY: Springer New York, 2014: 227-250.
- [94] <https://www.intuitive.com/en-us/products-and-services/ion.2019-12>
- [95] Parent F, Loranger S, Mandal K K, et al. Enhancement of accuracy in shape sensing of surgical needles using optical frequency domain reflectometry in optical fibers[J]. Biomedical Optics Express, 2017, 8(4): 2210-2221.
- [96] Parent F, Gerard M, Monet F, et al. Intra-arterial Image Guidance with Optical Frequency Domain Reflectometry Shape Sensing[J]. IEEE Transactions on Medical Imaging, 2019, 38(2): 482-492.
- [97] Davis M A, Malinowski E, Chevalier J L, et al. Body shape, position, and posture recognition suit with multi-core optical shape sensing fiber: US, US9304018[P].2015-05-07.
- [98] Jang M, Kim J S, Kang K, et al. Towards finger motion capture system using FBG sensors [C].2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE,2018:3734-3737.
- [99] Lalam N, Ng W P, Dai X W, et al. Performance analysis of Brillouin optical time domain reflectometry (BOTDR) employing wavelength diversity and passive depolarizer techniques[J]. Measurement Science and Technology, 2018, 29 (2): 1-9.
- [100] M. D. Jones. Using Simplex Codes to Improve OTDR Sensitivity[J]. IEEE Photonics Technol. Lett., 1993, 15(7): 822 - 824.
- [101] M. Taki, M. A. Soto, G. Bolognini, and F. Di Pasquale. Long-Range Distributed Strain and Temperature Sensing with 40-cm Spatial Resolution Based on DPP-BOTDA Employing Optical Pre-Amplification and Simplex Coding[C]Optical Fiber Communication Conference/National Fiber Optic Engineers Conference 2011 (OFC/NFOEC), March 6 – March 10, 2011, Los Angeles, California United States. OSA Technical Digest (CD) (Optica Publishing Group, 2011), paper OTuL2.
- [102] Feng Wang, Weiwei Zhan, Yuangang Lu, Zhijun Yan, Xuping Zhang. Determining the Change of Brillouin Frequency Shift by Using the Similarity Matching Method[J]. Journal of Lightwave Technology, 2015, 33(19): 4101-4108.
- [103] Chang Y Q, Wu H, Zhao C, et al. Distributed Brillouin frequency shift extraction via a convolutional neural network[J]. Photonics Research,2020, 8(5): 690-697.
- [104] Wang F, Zhan W W, Zhang X P, et al. Improvement of spatial resolution for BOTDR by iterative subdivision method[J]. Journal of Lightwave Technology, 2013, 31(23): 3663–3667.
- [105] Moffat R, Sotomayor J, Beltrn J F. Estimating tunnel wall displacements using a simple sensor based on a Brillouin optical time domain reflectometer apparatus[J]. International Journal of Rock Mechanics and Mining Sciences, 2015, 75: 233-243.
- [106] Cheng G, Shi B, Zhu H H, et al. A field study on distributed fiber optic deformation monitoring of overlying strata during coal mining[J]. Journal of Civil Structural Health Monitoring, 2015, 5(5): 553-562.
- [107] Antman Y, Clain A, London Y, et al. Optomechanical sensing of liquids outside standard fibers using forward stimulated Brillouin scattering[J]. Optica, 2016, 3(5): 510-516.

- [108] Bashan G, Diamandi H H, London Y, et al. Optomechanical time-domain reflectometry[J]. Nature Communications, 2018, 9(1): 2991.
- [109] Chow D M, Yang Z, Soto M A, et al. Distributed forward Brillouin sensor based on local light phase recovery[J]. Nature Communications, 2018, 9(1): 2990.
- [110] Ohashi M, Shibata N. Fiber diameter estimation based on guided acoustic wave Brillouin scattering[J]. Electronics Letters, 1992, 28(10):900-902.
- [111] Shiraki K, Ohashi M. Sound velocity measurement based on guided acoustic-wave Brillouin scattering[J]. Photonics Technology Letters IEEE, 1992, 4(10):1177-1180.
- [112] Tanaka Y, Ogusu K. Tensile-strain coefficient of resonance frequency of depolarized guided acoustic-wave Brillouin scattering[J]. IEEE Photonics Technology Letters, 1999, 11(7):865-867.
- [113] Hayashi N, Suzuki K, Set S Y, et al. Temperature coefficient of sideband frequency produced by polarized guided acoustic-wave Brillouin scattering in highly nonlinear fibers[J]. Appl. phys.express, 2017, 10(9):092501.
- [114] Zaixuan Zhang, Jianfeng Wang, Xiangdong Yu. Research on temperature measurement by Raman scattering distributed fiber[J]. Photoelectrons. Laser, 2001, 12(6): 596-600.



Building a Fully Connected,
Intelligent World

Hello, Greater Bay Area!

Huawei and China Mobile Guangdong have jointly built an all-optical network cluster for the Guangdong-Hong Kong-Macao Greater Bay Area. The network cluster has served over 110 million individuals, 18 million households, and 2 million enterprises, while reducing electricity use by over 10 million kWh and carbon emissions by over 6400 tons for local communication networks each year. This helps balance economic development and environmental protection.





Building a Fully Connected,
Intelligent World



OptiX for Good

OptiX for Good

Huawei OptiX for Good Plan



Building low-carbon, efficient, and intelligent all-optical networks, bridging the digital divide, and enabling all-optical networks to benefit every person

The background of the page features a futuristic cityscape with a glowing network overlay. The city is composed of numerous skyscrapers and buildings, rendered in shades of blue, white, and gold. A complex network of lines and nodes, resembling a data or communication network, is superimposed over the city and extends into the sky. The overall aesthetic is high-tech and digital.

HUAWEI TECHNOLOGIES CO., LTD.

Huawei Industrial Base
Bantian, Longgang
Shenzhen 518129, PRC
Tel: +86-755-28780808

Trademark Notice

 HUAWEI, HUAWEI, and  are trademarks or registered trademarks of Huawei Technologies Co., Ltd.
All other trademarks and product, service, and company names mentioned in this journal are the property of their respective owners.

General Disclaimer

The information in this journal may contain predictive statement including, without limitation, statements regarding the future financial and operating results, future product portfolios, and new technologies. There are a number of factors that could cause actual results and developments to differ materially from those expressed or implied in the predictive statements. Therefore, such information is provided for reference purpose only and constitutes neither an offer nor an acceptance. Huawei may change the information at any time without notice.

Copyright © 2023 Huawei Technologies Co., Ltd. All Rights Reserved.

No part of this journal may be reproduced or transmitted in any form or by any means without prior written consent of Huawei Technologies Co., Ltd.