



中文核心期刊 中国科技核心期刊 中国核心学术期刊
第三届国家期刊奖百种重点期刊 信息通信领域产学研合作特色期刊

ISSN 1009-6868
CN 34-1228/TN

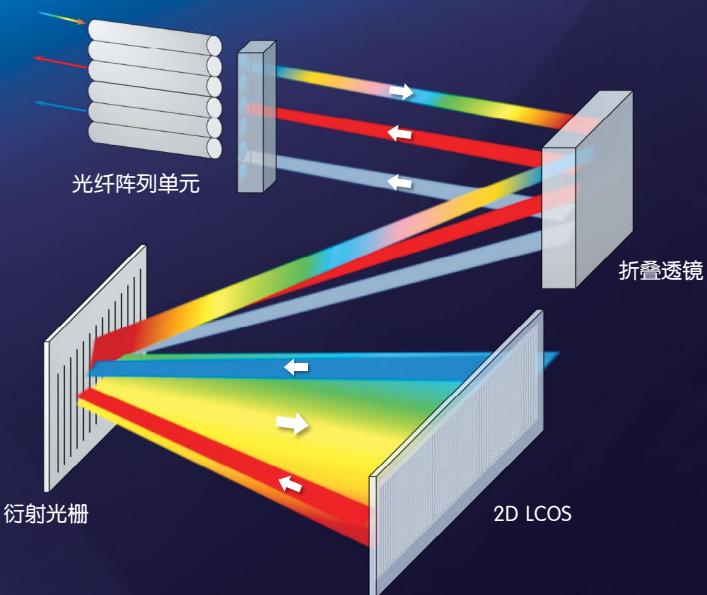
中兴通讯技术

ZTE TECHNOLOGY JOURNAL

<http://tech.zte.com.cn>

第 31 卷 · 总第 186 期 · 2025 年 12 月 · 第 6 期 (卷终)

专题：新一代光传输技术



(封面图片详解见 P46)

ISSN 1009-6868



《中兴通讯技术》第10届编辑委员会

顾问	侯为贵(中兴通讯股份有限公司创始人)	钟义信(北京邮电大学教授)
	糜正琨(南京邮电大学教授)	李自学(中兴通讯股份有限公司前董事长)
主任	陆建华(中国科学院院士)	
副主任	方榕(中兴通讯股份有限公司董事长) 李建东(西安电子科技大学教授)	

编 委

陈建平	上海交通大学教授	唐雄燕	中国联通研究院副院长
陈前斌	重庆邮电大学教授、副校长	陶小峰	北京邮电大学教授
段晓东	中国移动研究院副院长	汪烈军	新疆大学教授、副校长
方榕	中兴通讯股份有限公司董事长	王翔	中兴通讯股份有限公司高级副总裁
葛建华	西安电子科技大学教授	王文博	雄安空天信息研究院院长
管海兵	上海交通大学教授、副校长	王文东	北京邮电大学教授
郭庆	哈尔滨工业大学教授	王喜瑜	中兴通讯股份有限公司执行副总裁
洪伟	中国科学院院士、东南大学教授	王耀南	中国工程院院士、湖南大学教授
江涛	华中科技大学教授	王志勤	中国信息通信研究院副院长
蒋林涛	中国信息通信研究院科技委主任	卫国	中国科学技术大学教授
金石	东南大学教授、副校长	邬贺铨	中国工程院院士
李尔平	中国工程院外籍院士、浙江大学教授	吴春明	浙江大学教授
李红滨	北京大学教授	向际鹰	中兴通讯股份有限公司首席科学家
李厚强	中国科学技术大学教授	肖甫	南京邮电大学教授、副校长
李建东	西安电子科技大学教授	解冲锋	中国电信新一代信息通信专业首席专家
李乐民	中国工程院院士、电子科技大学教授	徐安士	北京大学教授
李融林	华南理工大学教授	徐子阳	中兴通讯股份有限公司总裁
林晓东	中兴通讯股份有限公司副总裁	续合元	中国信息通信研究院首席专家
刘健	中兴通讯股份有限公司高级副总裁	薛向阳	复旦大学教授
刘建伟	北京航空航天大学教授	杨义先	北京邮电大学教授
隆克平	北京科技大学教授	易芝玲	中国移动研究院首席科学家
卢光跃	西安邮电大学教授、校长	张杰	内蒙古工业大学教授、学术副校长
陆建华	中国科学院院士、清华大学教授	张平	中国工程院院士、北京邮电大学教授
马建国	中原工学院教授、学术副校长	张卫	复旦大学教授
毛军发	中国科学院院士、深圳大学校长	张宏科	中国工程院院士、北京交通大学教授
孟洛明	北京邮电大学教授	张钦宇	哈尔滨工业大学(深圳)教授、副校长
尼玛扎西	中国工程院院士、西藏大学教授	张云勇	中国联通网络信息安全部总经理
石光明	鹏城实验室副主任	赵慧玲	工业和信息化部信息通信科技委常委
史振威	内蒙古大学教授	郑纬民	中国工程院院士、清华大学教授
孙知信	南京邮电大学教授	钟章队	北京交通大学教授
谈振辉	北京交通大学教授	周亮	南京邮电大学教授、副校长
唐宏	中国电信IP专业首席专家	朱近康	中国科学技术大学教授
唐万斌	电子科技大学教授	祝宁华	中国科学院院士、南开大学教授

目 次

中兴通讯技术 (ZHONGXING TONGXUN JISHU)
第31卷 总第186期 2025年12月 第6期(卷终)

中文核心期刊 中国科技核心期刊 第三届国家期刊奖百种重点期刊 信息通信领域产学研合作特色期刊 中国知网、万方数据、重庆维普等数据库收录期刊 1995年创刊

热点专题 ▶

新一代光传输技术

- 01 专题导读 陈建平, 唐雄燕
03 基于低复杂度Transformer的光纤信道快速精确建模技术 史明辉, 郑智雄, 牛泽坤, 义理林
13 面向智算中心互联的光算协同技术研究 谭艳霞, 满祥锟, 吴绍辉, 张贺, 徐博华
20 50G-PON平滑演进的挑战和方案研究 杨波, 张德智, 刘波
29 数字副载波复用光纤通感融合系统干涉衰落抑制研究
..... 张兵兵, 向梦, 张泽彬, 付松年, 秦玉文
39 新一代光传输网络关键器件和集成化技术研究 罗勇, 张冀, 梁雪瑞
49 关于AI原生的几点探讨 何宝宏
53 智算光互联技术进展及趋势 张平化, 王会涛, 王勇
61 基于熵函数的语义信息边界研究 唐雪, 许进, 冯雨龙
70 智能超表面自动化部署解决方案 李萍, 叶峥嵘, 窦建武
75 面向6G的天地一体化网络安全技术 陆海涛, 周强, 代九龙, 卢帆, 李锐
I 《中兴通讯技术》第31卷总目次
III 《中兴通讯技术》2026年专题计划

名家视点 ▶

企业视界 ▶

技术广角 ▶

综合信息 ▶

《中兴通讯技术》2025年热点专题名称及策划人

1. 6G立体覆盖技术

西安电子科技大学教授 李建东
西安电子科技大学教授 刘俊宇

3. 6G网络安全

北京航空航天大学教授 刘建伟
北京航空航天大学教授 王景璟

5. 网络中的AI技术

中国电信研究院教授级高工 解冲锋
北京邮电大学教授 孟洛明
清华大学教授 崔勇

2. 智算网络

中国移动研究院副院长 段晓东
清华大学教授 李丹
电子科技大学教授 虞红芳

4. 面向6G的高时效智能机器通信技术

中国工程院院士、北京邮电大学教授 张平
北京邮电大学副教授 秦晓琦

6. 新一代光传输技术

上海交通大学教授 陈建平
中国联通研究院副院长 唐雄燕

MAIN CONTENTS

ZTE TECHNOLOGY JOURNAL
Vol. 31 No. 6 (End of Volume) Dec. 2025

Special Topic ►

Next-Generation Optical Transmission Technology

- 01 Editorial CHEN Jianping, TANG Xiongyan
- 03 Fast and Accurate Optical Fiber Channel Modeling Technology Based on Low-Complexity Transformer SHI Minghui, ZHENG Zhixiong, NIU Zekun, YI Lilin
- 13 Optical-Computing Collaborative Technology for Intelligent Computing Center Interconnection TAN Yanxia, MAN Xiangkun, WU Shaohui, ZHANG He, XU Bohua
- 20 Challenges and Solutions for Smooth Evolution of 50G-PON YANG Bo, ZHANG Dezhi, LIU Bo
- 29 Interference Fading Suppression in Fiber-Optic Integrated Sensing and Communication Systems Using Digital Subcarrier Multiplexing ZHANG Bingbing, XIANG Meng, ZHANG Zebin, FU Songnian, QIN Yuwen
- 39 Key Devices and Integration Technology of Next-Generation Optical Transmission Network LUO Yong, ZHANG Ji, LIANG Xuerui
- 49 Several Insights into AI Nativism HE Baohong
- 53 Advances and Evolutionary Trends in Optical Interconnection Technology for Intelligent Computing ZHANG Pinghua, WANG Huitao, WANG Yong
- 61 Research on Semantic Information Boundary Based on Entropy TANG Xue, XU Jin, FENG Yulong
- 70 Automatic Deployment Solution for Reconfigurable Intelligent Surface LI Ping, YE Zhengzheng, DOU Jianwu
- 75 Network Security Technology for Space-Ground Integrated Networks Towards 6G LU Haitao, ZHOU Qiang, DAI Jiulong, LU Fan, LI Rui

Expert View ►

Enterprise View ►

Research Papers ►

期刊基本参数：CN 34-1228/TN*1995*b*16*83*zh*P*¥20.00*6500*11* 2025-12

敬告读者

本刊享有所发表文章的版权，包括英文版、电子版、网络版和优先数字出版版权，所支付的稿酬已经包含上述各版本的费用。未经本刊许可，不得以任何形式全文转载本刊内容；如部分引用本刊内容，须注明该内容出自本刊。

新一代光传输技术专题导读



专题策划人



陈建平



唐雄燕

中国工程院首批外籍院士、前美国光学学会主席、美国国家工程院院士厉鼎毅先生，曾将光纤的发明誉为20世纪最伟大的技术发明。近40年来，光纤通信蓬勃发展，极大地推动了信息技术的广泛应用。当前，传统电信业务持续增长，以人工智能（AI）为代表的新业务呈现爆炸式增长态势。无论是骨干网、接入网，还是数据/算力中心乃至片上互联，都呼唤着新一代光传输技术，以应对在容量、时延、功耗和成本等方面挑战。此外，光纤在传感等领域的优势正推动通信与传感的一体化发展。为此，本期以“新一代光传输技术”为主题，特邀学术界与产业界的专家学者，深入探讨其中的关键技术问题，分享建设性的解决方案。

《基于低复杂度Transformer的光纤信道快速精确建模技术》一文，针对现有基于物理模型的分步傅里叶算法迭代运算量大、复杂度高的问题，提出了一种基于Transformer时序神经网络的建模架构。该方法将绝对位置编码优化为旋转位置编码，从而更有效地感知符号间相对位置信息，使其更贴合光纤非线性的物理特征，显著提升了建模精度；同时，将传统的全局自注意力机制改进为滑动窗口注意力机制，使每个符号仅需计算在其非线性关联长度范围内的相关性，从而

大幅降低了计算复杂度。光纤信道建模对于表征光纤特性、开发先进数字信号处理算法具有重要意义。该文提出的建模技术有望成为下一代高速光纤通信系统中信道建模的有效工具，进一步推动光通信技术的发展。

《面向智算中心互联的光算协同技术研究》一文，针对智算中心互联对光网络提出的新需求，系统剖析了其核心诉求与技术挑战，深入探讨了智算互联架构及光算协同关键技术，涵盖确定性承载、安全可靠、任务式带宽、长距离无损传输以及算网协同管控编排等方向。文章进一步介绍了智算网络的典型应用实践：面向跨智算中心分布式协同训练，搭建了基于光传送网（OTN）互联的现网试验环境，在广域收敛比不低于16:1的条件下，实现了百亿参数AI大模型跨域分布式训练性能不低于95%；试验采用单波800 Gbit/s速率，成功完成300 km传输验证，有力证明了光算协同技术在长距离、高可靠、大带宽算力互联场景中的实用价值。该研究不仅为智算中心互联提供了可落地的技术方案，也为运营商拓展算网一体化服务、创新商业模式提供了重要参考。

《50G-PON平滑演进的挑战和方案研究》一文，针对现网宽谱以太网无源光网络（EPON）终端存量较大的运营商，提出一种50G-PON三代时分共存方案，旨在解决从EPON、10G-EPON向50G-PON演进过程中面临的宽谱终端波长冲突与兼容性难题。论文研究了50G-PON光线路终端（OLT）

侧采用下行 $1366 \pm 2 \text{ nm}$ 负啁啾电吸收调制激光器 (EML) 与上行四速率时分复用突发接收机, 保障了 20 km 传输距离与 29 dB 功率预算; 并通过融合三模 Combo 光模块与统一动态带宽分配 (DBA) 调度, 实现了 EPON、10G-EPON 和 50G-PON 三代系统及 6 类终端的全兼容共存, 相关方案已在实验室环境中完成测试验证。该研究为运营商在现有 EPON 区域部署 50G-PON 设备, 提供了一种有效降低部署成本与实施难度的技术路径。

《数字副载波复用光纤通感融合系统干涉衰落抑制研究》聚焦通感融合 (ISAC) 这一信息光子技术重要方向, 针对基于数字副载波复用 (DSM) 架构的相位敏感光时域反射 (Φ -OTDR) 技术所面临的干涉衰落问题, 提出一种接收端多域集成的数字信号处理解决方案。该方案创新性地融合相位域、频率域与空间域的技术优势, 通过协同作用有效抑制干涉衰落, 且无须额外增加硬件复杂度, 兼顾了系统灵活性与成本效益。论文通过系统实验验证了该方案的可行性, 为提升通感融合系统的环境适应性与实用化水平提供了关键技术支撑, 也为光电子系统中的多域协同信号处理提供了新思路。

《新一代光传输网络关键器件和集成化技术研究》一文, 围绕新一代光传输网络向超大容量、超长距离、高灵活性和低能耗方向发展的需求, 系统梳理了关键器件与集成化技术的相关研究进展。文章综述了光电集成技术平台的特点与选择, 并深入探讨了可调谐光源、集成化收发芯片、集成光波导放大器 (EDWA)、智能光交换等关键器件及先进封装技术的最新发展。同时, 文章还分析了上述技术与器件在超大

容量传输系统、数据中心互连及 5G 承载网中的应用, 阐述了多材料异构集成、热管理和封装成本等方面面临的挑战, 并对共封装光学、晶圆级制造与智能化光网络等未来发展趋势进行了展望。

本期专题的作者来自知名高校、光通信头部企业与科研机构, 从器件、系统、通感算协同等角度, 对新一代光传输技术展开了深入剖析。期待这些高质量的研究成果能够为高校师生、光通信企业与通信运营商提供有益的参考和启示, 并在此对所有作者和审稿专家的大力支持表示由衷的感谢!

策划人简介

陈建平, 上海交通大学特聘教授, 曾任第七届国务院学位委员会学科评议组成员, 目前担任中国通信学会第一届光电融合技术专业委员会副主任委员; 长期从事光通信与光电子集成技术研究和教学工作; 作为首席科学家承担完成国家“973”计划项目“超高速低功耗光子信息处理集成芯片与技术基础研究”, 曾获国家教学成果奖二等奖, 策划创建的上海交大平湖智能光电研究院已成为中国光电子芯片封装、测试与失效分析重要平台; 作为副主编, 出版《光电集成技术》丛书。

唐雄燕, 教授级高工, 现任中国联通研究院副院长、首席科学家, 下一代互联网宽带业务应用国家工程研究中心主任, 新世纪百千万人才工程国家级人选, 北京邮电大学兼职教授、博士生导师, 工业和信息化部信息通信科技委委员, 中国通信学会理事和会士, 中国光学工程学会常务理事和会士; 长期从事信息通信新技术新业务研发和管理工作, 主要专业领域为宽带通信、光纤传输、互联网/物联网、新一代网络等; 发表论文 200 余篇, 出版专著 11 部。



基于低复杂度Transformer的光纤信道快速精确建模技术

Fast and Accurate Optical Fiber Channel Modeling Technology
Based on Low-Complexity Transformer

史明辉/SHI Minghui, 郑智雄/ZHENG Zhixiong,
牛泽坤/NIU Zekun, 义理林/YI Lilin
(上海交通大学, 中国 上海 200240)
(Shanghai Jiao Tong University, Shanghai 200240, China)

DOI: 10.12142/ZTETJ.202506002
网络出版地址: <https://link.cnki.net/urlid/34.1228.TN.20251230.1452.002>
网络出版日期: 2025-12-30
收稿日期: 2025-10-28

摘要: 光纤信道建模对于表征光纤特性、开发先进数字信号处理算法十分重要。基于物理模型的分步傅里叶算法 (SSFM) 需要大量迭代运算, 复杂度较高, 限制了其应用前景。提出了一种基于低复杂度 Transformer 架构的光纤信道建模方法。该方法对传统 Transformer 架构做了两处关键改进——采用相对位置编码替代绝对位置编码, 并采用滑动窗口注意力机制替代全局注意力机制, 从而提升模型对光纤信道非线性效应的表征能力。结果表明, 所提方法的有效信噪比 (ESNR) 与 SSFM 相比仅相差 0.15 dB, 计算时间较传统 Transformer 减少 69.9%, 较 SSFM 降低 96.9%, 从而在保持较高精度的同时, 显著降低了计算复杂度。

关键词: 光传输系统; 光纤信道建模; Transformer

Abstract: Optical fiber channel modeling is essential for characterizing fiber transmission characteristics and developing advanced digital signal processing algorithms. The split-step Fourier method (SSFM), as a physics-based numerical method, provides high accuracy but suffers from high computational complexity due to its iterative nature, which limits its practical deployment. A low-complexity Transformer-based architecture for modeling optical fiber channel waveforms is proposed. It incorporates two key modifications to the standard Transformer: replacing absolute positional encoding with relative positional encoding and substituting global attention with a sliding-window attention mechanism. These changes enhance the model's ability to capture nonlinear transmission effects. Results show that the proposed method achieves an Effective Signal-to-Noise Ratio (ESNR) within only 0.15 dB of that obtained by SSFM, while reducing the computation time by 69.9% compared to the traditional Transformer and by 96.9% compared to SSFM. This verifies that the method maintains high accuracy while significantly lowering computational complexity.

Keywords: optical transmission system; optical fiber channel modeling; Transformer

引用格式: 史明辉, 郑智雄, 牛泽坤, 等. 基于低复杂度Transformer的光纤信道快速精确建模技术 [J]. 中兴通讯技术, 2025, 31(6): 3-12. DOI: 10.12142/ZTETJ.202506002

Citation: SHI M H, ZHENG Z X, NIU Z K, et al. Fast and accurate optical fiber channel modeling technology based on low-complexity transformer [J]. ZTE technology journal, 2025, 31(6): 3-12. DOI: 10.12142/ZTETJ.202506002

光纤通信系统是现代信息传输基础设施的骨干。快速精确地建模光纤信道, 对于评估光纤特性、优化光网络配置^[1-5]、开发先进的数字信号处理 (DSP) 算法^[6-10]以及实现端到端 (E2E) 优化^[11-14]具有关键作用。光信号在光纤中的传播服从非线性薛定谔方程 (NLSE)^[15], 该方程在大多数情况下缺乏解析解。高斯噪声 (GN) 模型及其变体^[16-17]属于功率级建模方法, 将非线性效应等效为高斯噪声, 能够快速、准确地估算广义信噪比。然而, 此类 GN 模型无法提供详细的信号波形信息, 因而在 DSP 算法开发中的应用受

限, 尤其在对非线性补偿要求较高的场景中。分步傅里叶方法 (SSFM)^[18]是一种传统的波形级建模方法, 通过迭代求解 NLSE。虽然 SSFM 精度较高, 但其计算效率低, 且复杂度与信号带宽呈四次方关系^[18], 这限制了其在下一代高速光传输系统中的应用。

近年来, 基于神经网络 (NN) 的数据驱动方法被应用于实现快速、准确的波形级光纤信道建模。这些方法通过利用带标签的数据与部分先验知识来学习信号与信道特性^[19]。基于 NN 的方法在保持与 SSFM 相近精度的同时, 能显著降

低计算复杂度。生成对抗网络 (GAN)^[20]、傅里叶神经算子 (FNO)^[21]等架构已在单通道系统中展现出优异的建模效果。时序神经网络, 如双向长短期记忆网络 (BiLSTM)^[19]和 Transformer^[22], 凭借其循环结构和自注意力机制, 更擅长捕捉具有符号关联性的光纤非线性特征, 表现出更高的建模精度。通过引入物理先验知识, 一种结合分布式特征解耦 (FDD) 与 BiLSTM 的方法已成功应用于多通道波分复用 (WDM) 系统。进一步地, 将 Transformer 引入 FDD 架构^[23], 提升了原有 BiLSTM 的非线性建模能力。然而, 传统 Transformer 所采用的绝对位置编码^[24]难以有效处理符号间非线性串扰所需的相对位置关系。此外, 其自注意力机制的计算复杂度与输入符号长度呈平方关系, 在符号关联性较强的长序列高速系统中仍面临复杂度较高的挑战。

为解决上述问题, 本文提出一种结合旋转位置编码 (RoPE) 和滑动窗口注意力机制^[25]的低复杂度 Transformer 架构 (Slide-Roformer)。该架构将绝对位置编码优化为 RoPE, 从而能更有效地感知符号间的相对位置信息, 使其更符合光纤非线性的物理特征, 以提升建模精度。同时, 本文将传统的全局自注意力机制改进为滑动窗口注意力机制, 采用局部感知方式, 使每个符号仅需计算在其非线性关联长度范围内的相关性, 从而显著降低了计算复杂度^[26]。实验结果表明, 在 5 通道、140 GBaud、发射功率为 8.5 dBm 的传输场景下, Slide-Roformer 相比传统 Transformer 在波形建模精度上提升

了 75%, 计算时间减少了 69.9%。此外, Slide-Roformer 与传统 SSFM 之间的有效信噪比 (ESNR) 误差仅为 0.15 dB, 且计算时间可降低 96.9%。这些结果表明, 本文提出的 Slide-Roformer 有望成为下一代高速光纤通信系统中信道建模的有效工具, 推动光通信技术的进一步发展。

1 原理

1.1 基于 SSFM 的光传输系统架构

在本章中, 我们主要介绍光传输系统的整体架构。该架构为 NN 的训练提供了数据来源, 同时也作为对比建模精度与计算复杂度的参考基准。一个典型的基于 SSFM 的光纤传输仿真系统如图 1 所示, 主要包括发射端、光纤信道和接收端。在发射端, 首先通过伪随机数种子生成比特序列, 并利用双偏振 16 阶正交幅度调制将其映射为符号序列。该符号序列随后以 4 倍过采样率进行上采样, 并通过滚降系数为 0.1 的根升余弦滤波器完成脉冲成形。接着, 各信道信号被调制到不同的载波频率上, 合成为全场的 WDM 信号, 并送入光纤信道进行传输。该 WDM 信号可表示为:

$$\mathbf{A}(z, t) = \sum_{k=1}^C \mathbf{A}_k(z, t) \exp(j\Delta\omega_k t) \quad (1),$$

其中, \mathbf{A} 表示在两种任意正交偏振模式下的光信号, 即 \mathbf{A}_x 和 \mathbf{A}_y 。 \mathbf{A}_k 表示第 k 个信道的光信号。变量 z 表示传输距离, t 表

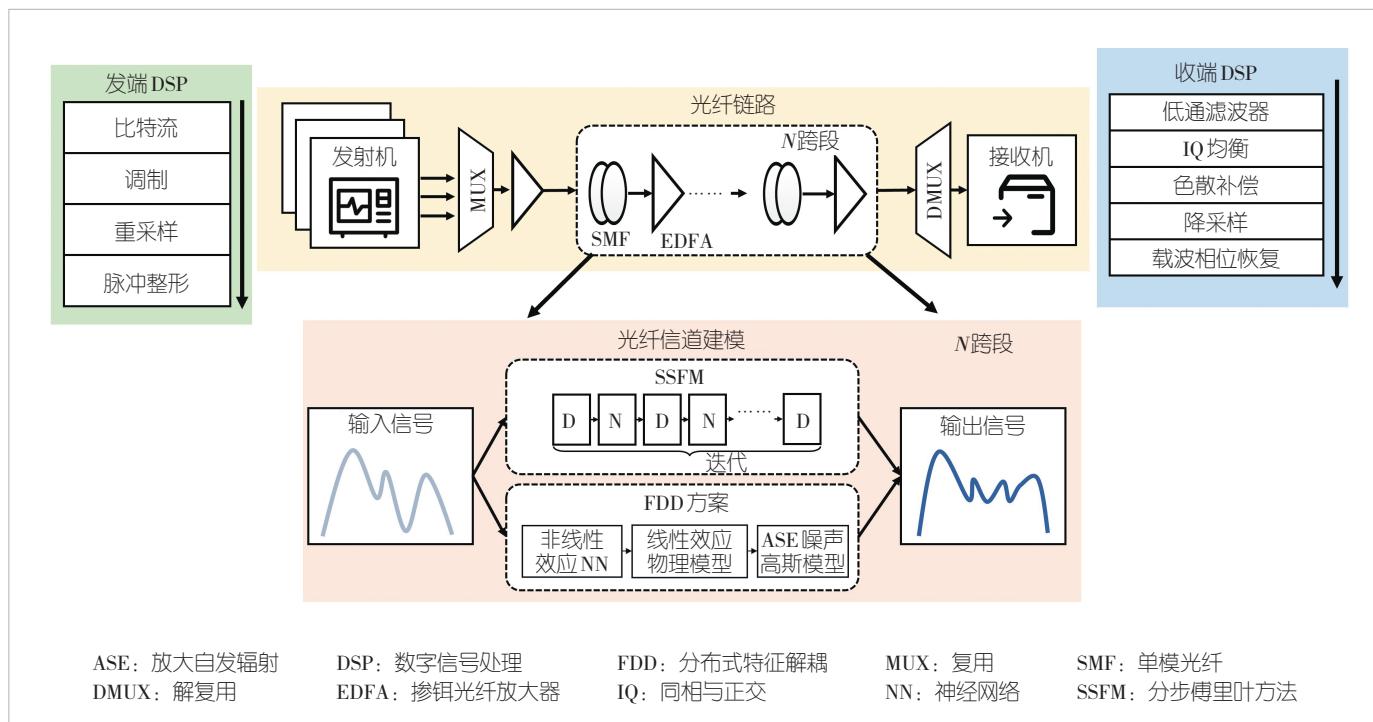


图1 光传输系统架构及信道建模方法

示时间坐标, C 为 WDM 通道总数, $\Delta\omega_k = \omega_k - \omega_0$ 为第 k 个信道中心频率与 WDM 信号中心频率之间的差值。

光信号在单模光纤中的传播由 NLSE 描述, 其形式为:

$$\frac{\partial \mathbf{A}(z, t)}{\partial z} = (\hat{\mathbf{D}} + \hat{\mathbf{N}}) \mathbf{A} \quad (2),$$

其中, $\hat{\mathbf{D}}$ 为线性算子, 用于描述衰减、色散 (CD); $\hat{\mathbf{N}}$ 为非线性算子, 用于描述克尔非线性效应, 包括自相位调制、交叉相位调制以及四波混频。在处理双偏振信号时, NLSE^[15] 通常采用耦合形式, 即耦合 NLSE, 或其简化形式——马纳可夫 (Manakov) ^[27-28] 方程。本文采用 Manakov 方程对光纤信道建模, 其形式为:

$$i \frac{\partial \mathbf{A}}{\partial z} - \frac{1}{2} \beta_2 \frac{\partial^2 \mathbf{A}}{\partial t^2} + \frac{8}{9} \gamma |\mathbf{A}|^2 \mathbf{A} + \frac{\alpha}{2} i \mathbf{A} = 0 \quad (3),$$

其中, β_2 为群速度色散参数, γ 为非线性参数, α 为损耗参数。CD 会引起符号间干扰 (ISI)。在 WDM 系统中, 受 ISI 影响的符号个数通过公式 (4) 表示:

$$N_{\text{ISI}} = \frac{\Delta T}{\Delta t} = \frac{I \beta_2 \Delta w}{1/S} = I \beta_2 \Delta w S \quad (4),$$

其中, N_{ISI} 表示受 ISI 影响的符号数, ΔT 是受 ISI 影响的时间宽度, Δt 是一个符号的持续时间, I 是传输距离, Δw 是频谱宽度, S 是符号速率。非线性与线性效应在光纤传输过程中同时存在并相互耦合, 这使得非线性也具备一定的符号关联性。非线性与信号的能量相关, 会在光纤的传输过程不断减弱, 导致其符号关联性比 CD 造成的 ISI 要弱。非线性的符号关联性可以采用有效非线性长度来近似:

$$L_{\text{eff}} = \frac{1}{\alpha}, \quad N_{\text{NL}} = L_{\text{eff}} \beta_2 \Delta w S \quad (5),$$

其中, L_{eff} 表示非线性有效长度, 通常小于实际传输长度。 N_{NL} 表示非线性符号间相关长度。对于当前时刻符号来说, 与之更加相近的符号对其造成的非线性串扰更强, 而更远端符号造成的影响则较弱。非线性强度与符号之间的相对位置关系有关, 而非绝对位置。

SSFM^[29-30] 是求解 NLSE 最常用的数据方法。该方法将长距离光纤传输路径划分为大量微小步长, 使得线性与非线性算子可独立处理。每一步的对称 SSFM 运算可表示为:

$$\mathbf{A}(z + h, t) \approx \exp\left(\frac{h}{2} \hat{\mathbf{D}}\right) \exp\left\{h \hat{\mathbf{N}} \left[\mathbf{A}\left(z + \frac{h}{2}, t\right)\right]\right\} \times \exp\left(\frac{h}{2} \hat{\mathbf{D}}\right) \quad (6),$$

其中, h 为单步长度。步长配置会影响 SSFM 的精度和计算效率。在本文中步长采用适用于非线性效应主导系统的变步长方法——非线性相位旋转法^[31]。采样率设置为单通道符号率

的 4 倍。为便于计算, 线性算子在频域中计算, 而非线性算子则在时域中计算。在每段光纤末端使用掺铒光纤放大器 (EDFA) 补偿信号衰减, 其引入的放大自发辐射 (ASE) 噪声可建模为高斯噪声。本文采用的光纤信道参数如表 1 所示。

在光纤信道传输后, 首先对目标信道信号进行解复用。接收端数字信号处理 (Rx DSP) 流程如下: 先使用匹配的根升余弦 (RRC) 滤波器进行滤波, 随后执行降采样。接着进行色散补偿 (CDC) 以校正线性损伤, 继而完成载波相位恢复 (CPR) 与解调。最终, 通过计算 Q 因子或有效信噪比 (ESNR) 等指标对传输性能进行定量评估。

基于 SSFM 的仿真系统所产生的 WDM 信号波形, 为 NN 模型提供了充足的训练数据。在性能测试阶段, 发送端与接收端采用完全一致的处理流程与参数设置, 以便公平比较基于 SSFM 与基于 NN 的两类光纤信道模型之间的性能差异。

1.2 基于分布式特征解耦的光纤信道建模

基于 SSFM 的光纤信道建模方法虽然精度较高, 但其计算复杂度高, 尤其在高波特率场景下, 限制了在下一代高速光传输系统中的应用。为缓解此问题, 研究人员引入 NN 进行光纤信道建模, 并结合部分物理先验信息, 构建数据-物理混合驱动的建模架构, 以进一步提升建模精度。FDD^[32] 是一种数据-物理混合驱动方案, 它通过物理模型建模线性效应、NN 建模非线性效应, 充分发挥二者优势, 从而提高整体建模精度。此外, FDD 每次仅建模一个光纤跨段, 通过多模块级联实现长距离传输。该方案在多通道 WDM 系统中展现出较好的建模效果, 因此本文选取该方案作为基础架构, 并对其中的 NN 结构进行优化。

为使 NN 有效学习具备符号关联性的非线性特征, 需通过滑动窗口方法构建包含过去与未来时刻填充符号的输入窗口。但在高速系统中, 由于符号关联性较长, 每次推理过程中需反复计算填充符号信息, 这导致复杂度较高。为此, 我们在 NN 的训练与推理过程中引入了序列到序列 (Seq2Seq) 架构^[30]。该架构采用多符号输入、多符号输出的模式, 能够

表 1 光纤信道参数

参数	数值	参数	数值
载波波长	1 550 nm	色散系数	17 ps/nm·km
通道数	5	非线性系数	1.3 / (W·km)
符号速率	140 GBaud	跨段长度	80 km
发射功率	8.5 dBm	EDFA 的噪声系数	5 dB
衰减	0.2 dB/km	SSFM 的最大非线性相位旋转	0.005

EDFA: 掺铒光纤放大器 SSFM: 分步傅里叶方法

单次同时预测多个符号，从而减少NN在重复循环预测过程中对填充符号的重复计算次数，降低推理时间。

此外，在训练过程中，参考“近端符号对当前时刻符号非线性串扰更强，远端符号更弱”的物理机制，我们采用迁移学习构建了两阶段训练模式。第一阶段训练中，先以较少的填充符号构建输入-输出数据对，以较低复杂度有效学习近端影响较大的非线性串扰；第二阶段训练中，进一步构建更长的输入-输出数据对，补充学习远端符号的非线性串扰。该两阶段训练模式有效降低了训练资源开销，并提升了模型的最终训练精度。

2 基于Slide-Roformer的光纤信道建模技术

BiLSTM凭借其序列建模能力，已成为光纤信道非线性建模的有效架构之一。Transformer基于其自注意力机制，相比BiLSTM具有更强的非线性拟合能力，在非线性更强的多通道、高功率WDM系统中表现出更优的建模性能。然而，传统Transformer所采用的绝对位置编码难以有效处理符号间非线性串扰所依赖的相对位置关系。此外，其自注意力机制的计算复杂度与输入符号长度呈平方关系，导致在符号关联性较长的高速系统中面临较高的复杂度挑战。在本文中，我们通过优化Transformer的位置编码与注意力机制，提出了Slide-Roformer架构，如图2所示，从而增强了Transformer在高速光通信系统中的信道建模能力。下文将分别介绍传统Transformer架构，以及优化后的旋转位置编码（RoPE）和滑动窗口注意力机制。

2.1 Transformer架构

光纤信道建模通常仅采用Transformer的编码器架构，而不使用解码器。这是因为仅凭编码器中的自注意力机制，就能实现对时序非线性特征的有效建模。Transformer编码器由多个堆叠的相同层组成，每层主要包括多头自注意力机制、前馈神经网络与残差连接3个核心模块。首先，多头自注意力机制通过并行使用多个不同的自注意力函数，从多个子空间中捕获不同类型的关系。随后，自注意力机制对输入信号进行3次不同的线性变换，分别生成查询（ Q ）、键（ K ）和值（ V ）向量，并通过公式（7）计算不同输入元素之间的关联程度。接着，多头注意力机制的输出经过前馈网络处理，以增强模型的表达能力。此外，每个子层之间均包含残差连接，这有助于缓解深层网络中的梯度消失问题，并提升信息流动效率。通过多层编码器的堆叠，模型能够逐步建立从局部到全局的多层次特征表示，从而实现有效的时序特征建模。

$$\text{Attention}(Q, K, V) = \text{Softmax}\left(\frac{QK^T}{\sqrt{d}}\right)V \quad (7)$$

此外，Transformer不具备类似BiLSTM的循环结构，无法有效获取不同元素间的位置关系。为了解决这一问题，Transformer通常引入绝对位置编码，通过一组确定性的正弦和余弦函数生成，其不同频率对应不同的时间尺度，从而在连续空间中为每个位置的元素分配唯一的位置表示形式。对于输入序列中第 pos 个位置和隐藏维度中的第 i 个维度，位

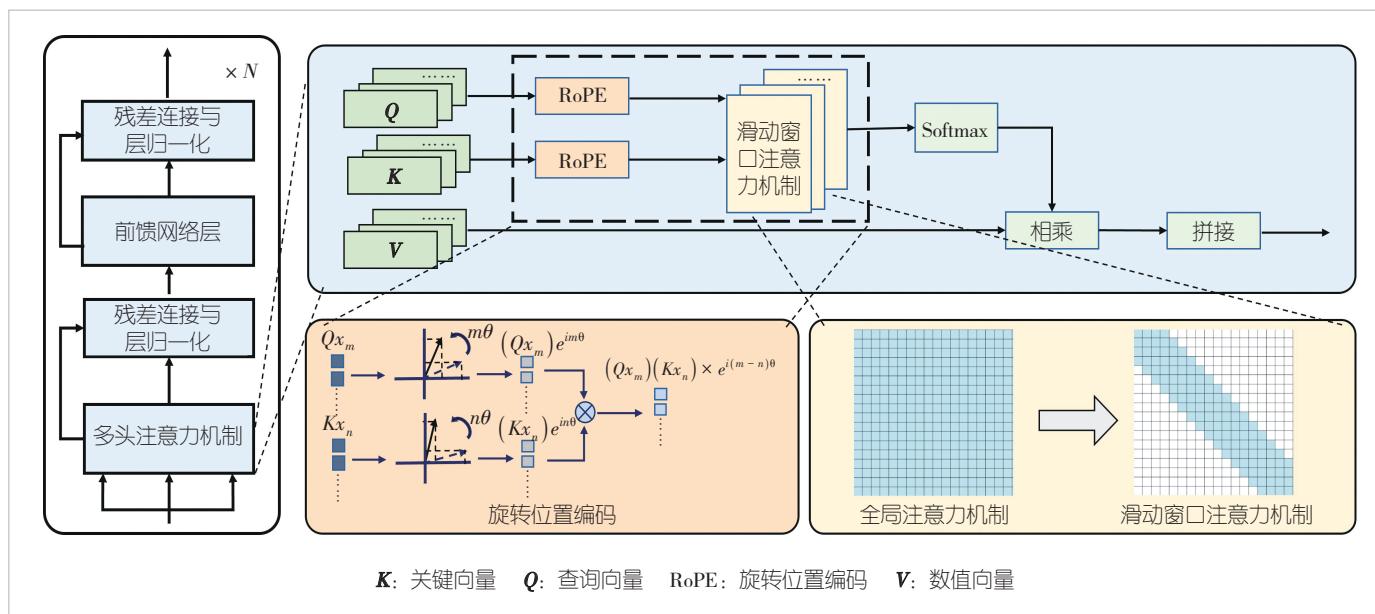


图2 Slide-Roformer结构

置编码的计算公式为：

$$\text{PE}(pos, 2i) = \sin\left(\frac{pos}{\theta^{\frac{2i}{d}}}\right), \text{PE}(pos, 2i + 1) = \cos\left(\frac{pos}{\theta^{\frac{2i}{d}}}\right) \quad (8),$$

其中， d 表示模型隐藏层的维度， θ 代表旋转角度。通过将位置编码与输入嵌入相加，模型在保持输入维度不变的情况下能够显式获得序列的顺序信息，从而在注意力计算中兼顾内容与位置信息。

2.2 旋转位置编码

绝对位置编码可以有效为输入元素添加绝对位置信息，但光纤信道的非线性效应主要取决于各符号之间的相对位置，而非绝对位置，这导致绝对位置编码方式效率低下，影响建模精度。为此，我们引入了一种相对位置编码方式—RoPE。该方法将位置信息以旋转变换的方式编码进各个位置的符号所对应的 Q 和 K 矩阵中，并通过 Q 和 K 矩阵相乘，以差分的形式表征各符号之间的相对位置关系，而不再直接将绝对位置信息添加到各符号中。RoPE的编码过程如图2金色部分所示，在位置 m 和 n 时， Q 和 K 分别经过公式(9)和公式(10)表示的旋转变换。其中， W_q 代表向量经过 Q 矩阵映射， W_k 代表经过 K 矩阵映射，并对 K 进行转置，再与 Q 相乘取实部，得到公式(11)。在相乘的结果中，通过 $m-n$ 差分地表示两个不同位置之间的相对位置关系。通过引入RoPE，模型可以更加高效地捕捉不同符号之间的相对非线性串扰强度，更加适配非线性串扰的物理特征，提高非线性建模精度。除此之外，引入RoPE的模型还具备较好的外推性，适配前文所提到的Seq2Seq方案。外推性是指模型在训练时和预测时输入长度不一致。这使得我们在模型的训练中采用输入长度相对较短序列，提高训练效率；在推理过程中，直接外推到更长的输入和输出序列，从而提高模型在预测长序列信号时的计算效率。

$$f_Q(\mathbf{x}_m, m) = (W_q \mathbf{x}_m) e^{im\theta} \quad (9),$$

$$f_K(\mathbf{x}_n, n) = (W_k \mathbf{x}_n) e^{in\theta} \quad (10),$$

$$g(\mathbf{x}_m, \mathbf{x}_n, m - n) = \text{Re} \left[(W_q \mathbf{x}_m) (W_k \mathbf{x}_n)^* e^{i(m-n)\theta} \right] \quad (11).$$

2.3 滑动窗口注意力机制

在光纤信道建模中，传统Transformer不仅存在位置编码上的不足，还受限于较高的计算复杂度。这一限制主要源于自注意力机制需计算当前符号与序列中所有符号之间的相对

关联，导致其复杂度与输入符号长度呈平方关系。此外，光纤的非线性符号关联性可由公式(5)表示，其关联长度被限制在一定范围内，超出此范围的信息对于当前符号而言是冗余的。上述问题使得模型在处理较长符号序列时产生大量冗余计算，增加了计算复杂度，尤其对符号关联性更长的下一代高速光传输系统影响显著。

为克服这一限制的约束，构建更加符合光纤信道非线性特征的注意力标准机制，我们引入了滑动窗口注意力机制，如图2黄色部分所示。在滑动窗口注意力中，对于每个位置 i 处的符号，仅计算其与当前位置前后 w 个符号的注意力，从而使得整个注意力窗口的大小 $2w + 1$ ，这显著降低了每个符号的计算复杂度。滑动窗口注意力机制首先使用一个滑动窗口操作在 K 、 Q 、 V 矩阵上提取局部窗口得到 K_{windowed} 、 Q_{windowed} 、 V_{windowed} 。我们以 K 的操作为例，滑动窗口过程可以表示为：

$$K_{\text{windowed}} \in R^{n \times (2w + 1) \times d} \quad (12).$$

每个 $K_{\text{windowed}}[i]$ 对应 K_{i-w}, \dots, K_{i+w} ，超出边界的可补零或掩码。随后使用张量广播方式，将每个 $Q_{i-w:i+w}$ 与 $K_{i-w:i+w}$ 进行了点积，最后得到使用滑动窗口注意力机制计算出的分数，如公式(13)所示：

$$\text{score}_i = \text{Softmax} \left(\frac{1}{\sqrt{d}} Q_{i-w:i+w} \cdot K_{i-w:i+w}^T \right) V_{i-w:i+w} \in R^{2w+1} \quad (13).$$

滑动窗口注意力机制相比全局注意力更高效，能在保持模型性能的同时，极大降低计算资源消耗，并支持长序列的高效建模。

3 结果分析

3.1 训练数据集和超参数配置

为验证Slide-Roformer相比传统Transformer(Vanilla-Transformer)在光纤信道建模中精度与复杂度的优势，我们构建了一套高速WDM相干光传输仿真系统。系统设置通道数为5，符号速率为140 GBaud，每通道发射功率为8.5 dBm，工作于高非线性区域，以考察两模型在强非线性条件下的建模能力差异。

训练过程中，为避免过拟合，我们采用10组不同随机种子生成长度为10 000符号的发送信号，并收集其经信道传输后各跨段的输入与输出波形作为训练数据集。Slide-Roformer与传统Transformer的参数配置如表2所示，二者主干网络参数基本一致。此外，为保证神经网络准确学习非线

表2 Slide-Roformer和Vanilla-Transformer的参数配置

参数	数值
输入尺寸	80
隐藏尺寸	240
注意力头数	6
前馈网络尺寸	960
层数	3

性效应,根据公式(5)计算非线性效应影响长度,将Slide-Roformer中滑动窗口注意力机制的长度 w 设为300。旋转角度 θ 为工程默认值通常设置为10 000,并使得最大的波长远大于常见序列长度,从而覆盖需要的频谱^[22, 25]。训练使用Smooth L1损失函数,优化器选用Adam,初始学习率为 5×10^{-4} ,并采用余弦退火策略逐步降低学习率。在Seq2Seq架构的两阶段训练中,第一阶段训练轮数设为500,输入窗口包含60个符号(中心20个为有效符号,前后各填充20个);第二阶段训练轮数为200,输入窗口包含580个符号(中心300个为有效符号,前后各填充140个)。

在测试阶段,我们采用与训练集不同的随机种子,生成长度为 1×10^6 符号的发送序列,分别通过不同模型进行传输,以比较其建模效果与计算时间。

3.2 性能标准

基于NN的光纤信道波形建模性能评估包含精度评估与复杂度评估两方面。精度评估基于波形建模精度和传输性能预测精度,并以SSFM的计算结果作为参考值。波形建模精度采用(NMSE)进行量化,其定义为:

$$NMSE = \frac{\sum_{i=1}^{N_{data}} |\hat{y}_i - y_i|^2}{\sum_{i=1}^{N_{data}} |y_i|^2} \quad (14),$$

y_i 表示SSFM的输出样本, \hat{y}_i 表示NN的输出样本。传输性能的评估需要先对建模信号进行Rx DSP处理并进行解调,计算传输的误码率(BER)和Q因子,它们之间的关系为:

$$Q = 20 \log_{10} \left(\sqrt{2} \operatorname{erfc}^{-1}(2\text{BER}) \right) \quad (15).$$

我们采用SSFM与NN所得的Q因子之间的误差,作为传输性能预测精度的度量指标。

复杂度评估则通过比较SSFM与NN在传输相同信号时所需的计算时间来完成。NN对于SSFM的加速比可表示为:

$$\text{Ratio} = \frac{\text{Time}_{\text{SSFM}} - \text{Time}_{\text{NN}}}{\text{Time}_{\text{SSFM}}} \quad (16).$$

3.3 Slide-Roformer的精度增益

为验证Slide-Roformer在建模精度上的提升,我们将其与Vanilla-Transformer进行对比。为进一步评估两阶段训练带来的性能增益,Slide-Roformer分别采用单阶段与两阶段训练进行实验。3个模型在训练过程中的损失函数变化如图3所示。

结果显示,采用两阶段训练的Slide-Roformer最终损失为 3.2×10^{-5} ,较单阶段训练的 5.9×10^{-5} 明显降低,这证明了两阶段训练对提升模型精度的有效性。进一步观察发现,在第一阶段训练结束时(第500轮),Slide-Roformer与Vanilla-Transformer的损失均约为 7.0×10^{-5} 。进入第二阶段后,Slide-Roformer的损失迅速收敛至 4.0×10^{-5} ,而Vanilla-Transformer的损失不仅未下降,反而上升至 2.5×10^{-2} ,凸显了Slide-Roformer在两阶段训练中的优势。

这一差异主要源于位置编码机制的不同:Slide-Roformer采用相对位置编码,即使两个阶段的输入符号长度不同,仍能通过相对位置关系有效提取相同的非线性特征;而Vanilla-Transformer采用绝对位置编码,当输入序列长度增加时,会引入第一阶段未见过的新位置信息,导致模型初始阶段无法适应此类信息,需重新训练收敛,从而影响了最终精度。

为进一步评估Slide-Roformer相对于传统方法的精度优

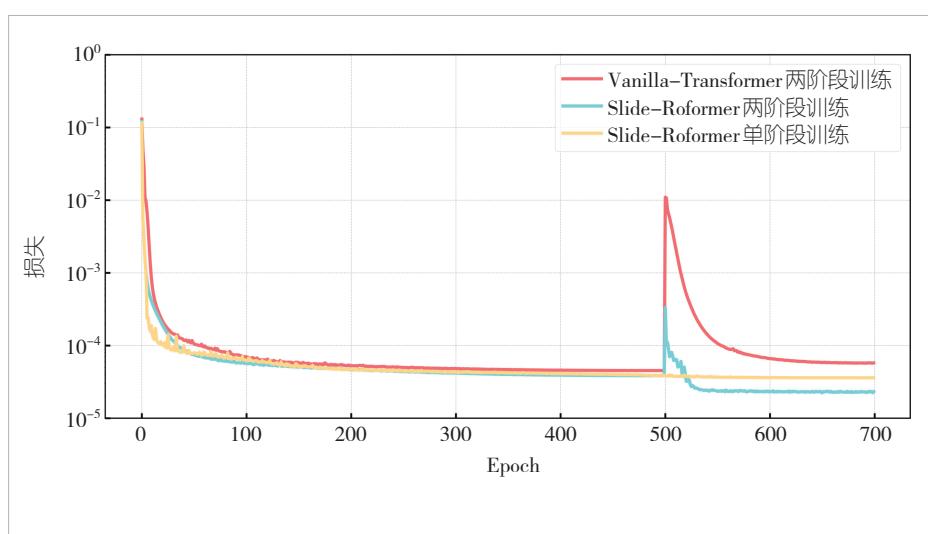


图3 Slide-Roformer与Transformer训练损失

势，我们将其与包括 Vanilla-Transformer、Seq2Seq-BiLSTM^[30]和 FNO^[21]在内的其他神经网络模型进行了对比。在 Seq2Seq-BiLSTM 中，BiLSTM 的隐藏层维度设为 240，层数为 3。该模型同样采用 Seq2Seq 架构，其两阶段训练的参数配置与 Slide-Roformer 保持一致：第 1 阶段输入窗口包含 60 个符号 (20+20+20)，第 2 阶段输入窗口包含 580 个符号 (140+300+140)。FNO 模型的傅里叶模态数设为 24，宽度为 16，层数为 4，采用单阶段训练，输入窗口包含 300 个符号 (140+20+140)。我们比较了各模型在 800 km 传输后的 NMSE，以评估其在长距离传输下的建模性能。如图 4 所示，Slide-Roformer 在 800 km 后的 NMSE 最低，仅为 2.0×10^{-3} ；而 Vanilla-Transformer、Seq2Seq-BiLSTM 和 FNO 的 NMSE 分别上升至 9.2×10^{-3} 、 4.2×10^{-3} 和 7.8×10^{-3} 。Slide-Roformer 相对于其他神经网络模型实现了约 50%~74% 的 NMSE 降低，这表明其具有更优的非线性效应建模能力。

其次，我们比较了深度学习模型与基准 SSFM 在传输性能预测上的差异，以评估它们在传输性能预测及 DSP 算法设计中的应用潜力。图 5 展示了基于 SSFM、Slide-Roformer 和 Vanilla-Transformer 建模的信号，在 800 km 传输并经过线性 DSP 处理后的星座图。可以看出，Slide-Roformer 与 SSFM 的星座图噪声分布高度吻合，而 Vanilla-Transformer 未能准确建模光纤信道的非线性噪声。从量化结果看，SSFM 的 ESNR 为 8.0 dB，Slide-Roformer 为 7.85 dB，误差仅为 0.15 dB；Vanilla-Transformer 的 ESNR 为 12.69 dB，与 SSFM 的误差达 4.69 dB。这些结果表明，Slide-Roformer 相比其他模型能更准确地反映光纤信道的特征。这些优势得益于 Slide-Roformer 采用的相对位置编码更符合光纤信道的非线性特性，且符号间的相对位置关系也使其能够适配更高效的两阶段训练模式，从而实现了更高的建模精度。

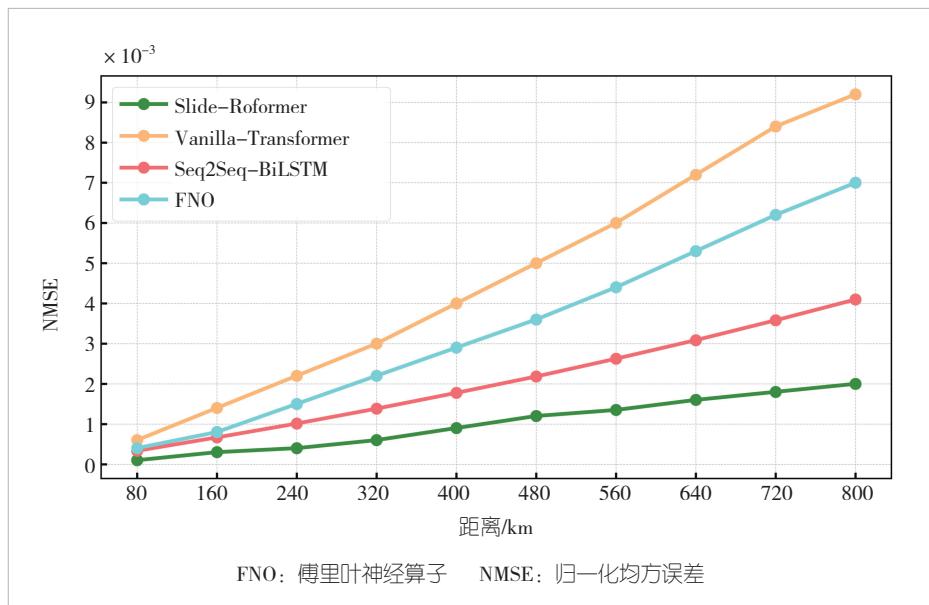


图 4 Slide-Roformer 与 Vanilla-Transformer 在 800 km 传输范围内 NMSE 的对比

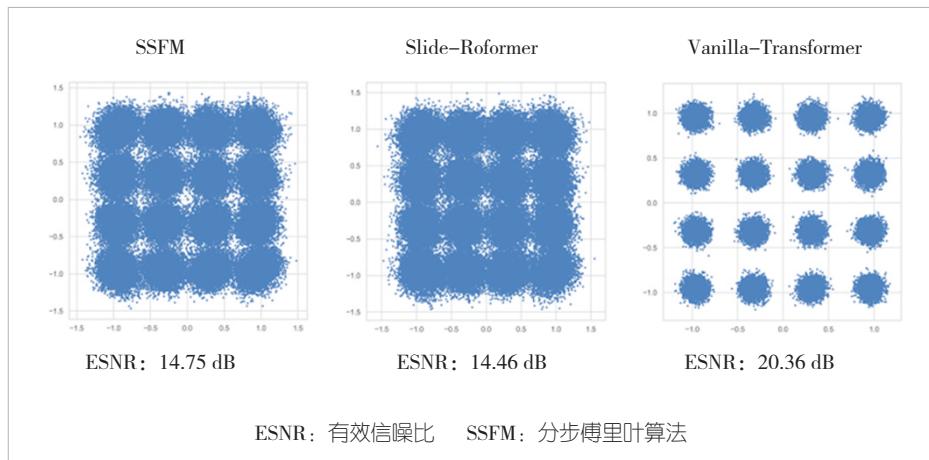


图 5 SSFM、Slide-Roformer、Vanilla-Transformer 在 800 km 处的星座图和对应的 ESNR

3.4 Slide-Roformer 复杂度优化

在模型推理过程中，采用多符号输出方式可有效减少两端填充符号的重复计算次数，从而缩短推理时间。为验证多符号输出机制的有效性，我们在建模相同总长度的符号序列时，分别设定模型单次推理预测 50 至 5 000 个符号，并记录完整信号建模所需时间。所有神经网络模型均部署于同一台配备 NVIDIA GeForce RTX 4090 24 GB GPU 的服务器上。

具体结果如图 6 所示，随着单次输出符号数量的增加，总计算时间逐渐减少，从输出 50 个符号时的 65 s 优化至输出 5 000 个符号时的 4.6 s，证明了该方法在降低计算时间方面的有效性。进一步地，为确保多符号输出模式下模型精度不发生劣化，我们测试了 Slide-Roformer 与 Vanilla-

Transformer在单次输出300~5 000个符号时的NMSE,如图7所示。

在800 km传输距离下,Slide-Roformer在不同输出符号长度下的NMSE整体稳定在约 2.8×10^{-3} 的水平,而Vanilla-Transformer的NMSE随输出符号长度的增加持续恶化,在单次输出5 000个符号时达到1.9,比Slide-Roformer高出3个数量级。这表明Slide-Roformer能够有效适用于多符号预测模式,而Vanilla-Transformer在多符号预测中出现了显著的波形失真,无法保持建模精度。

这一优势得益于Slide-Roformer在位置编码与注意力机制上的有效设计:首先,RoPE采用的相对位置编码保证了模型在不同输入长度下仍能有效感知符号间的相对位置信息;其次,滑动窗口注意力机制将符号的关联范围限制在特定长度内,使得无论输入序列多长,每个时刻的符号均可在固定范围内以相同模式学习非线性相关性,从而实现稳定有效的非线性建模。

为进一步评估计算效率,我们将Slide-Roformer与数值方法SSFM以及神经网络模型Vanilla-Transformer、Seq2Seq-BiLSTM和FNO进行了计算时间对比。Slide-Roformer与Seq2Seq-BiLSTM均单次预测5 000个符号;而Vanilla-Transformer为了保持较高精度,单次预测300个符号;FNO则与训练配置一致,单次预测20个符号。表3统计了上述方法在传输80 km(单个跨段)时的计算时间。结果表明,Slide-Roformer的计算时间仅为4.6 s,较Vanilla-Transformer(15.3 s)、Seq2Seq-BiLSTM(10.9 s)和FNO(91.2 s)具有明显优势。FNO因采用少符号预测模式,需多次循环并重复计算填充符号,导致复杂度较高。相比之下,Slide-Roformer

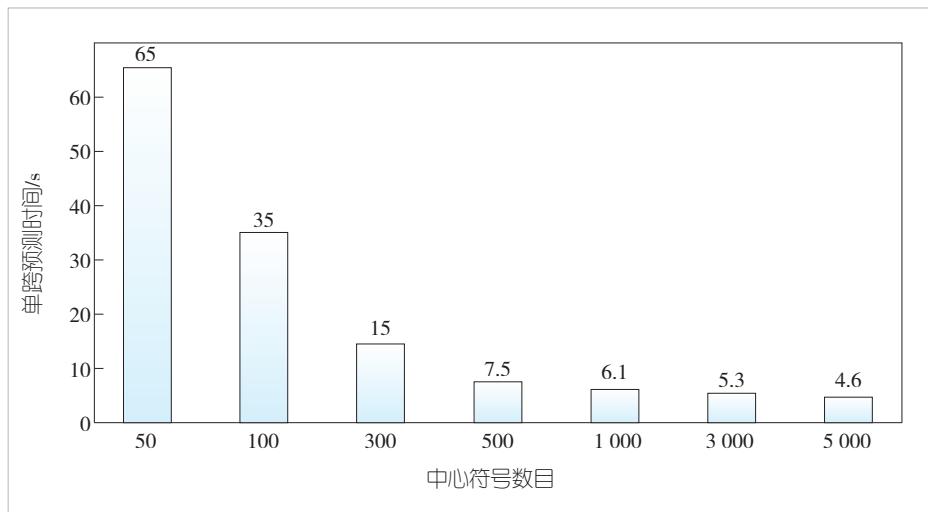


图6 不同输出中心符号长度下神经网络的单跨预测时间

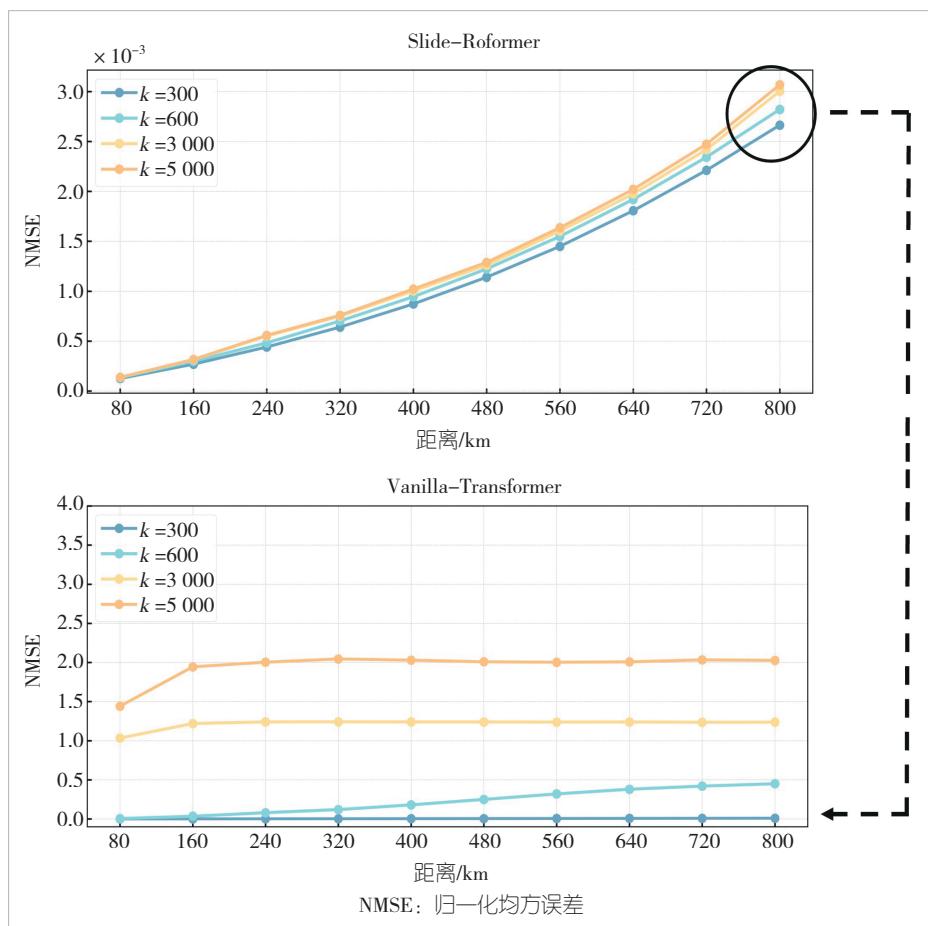


图7 Slide-Roformer和Transformer输出不同符号的NMSE对比

表3 不同模型的计算时间(传输距离80 km)

模型	Slide-Roformer	Vanilla-Transformer	Seq2Seq-BiLSTM	FNO	SSFM
时间/s	4.6	15.3	10.9	91.2	151

FNO: 傅里叶神经算子 SSFM: 分步傅里叶算法

通过将全局注意力优化为滑动窗口注意力，在多符号预测模式下降低了注意力模块的计算复杂度，从而较传统Transformer更为高效。与传统数值方法SSFM（仿真时间151 s）相比，Slide-Roformer实现了96.9%的加速。以上结果充分表明，Slide-Roformer能够有效克服传统算法复杂度高的难题，有望成为下一代高速光传输系统中信道建模的高效工具。

4 结语

本文针对传统分步傅里叶方法计算复杂度高的问题，提出了一种用于光纤信道建模的低复杂度Slide-Roformer架构。通过引入RoPE与滑动窗口注意力机制，该模型能够更有效地学习符号间的相对非线性串扰关系，并显著减少计算时间。实验结果表明，与传统Transformer相比，Slide-Roformer的波形误差NMSE改善了74%，计算时间降低了69.9%。与SSFM方法相比，ESNR仅为0.15 dB，计算时间可降低96.9%。这些结果验证了Slide-Roformer在光纤信道建模中兼具高精度与低复杂度的优势，为相关研究提供了一种高效的新方法，有望推动下一代高速光通信系统的进一步发展。

参考文献

- [1] ZHANG Y, PANG X H, SONG Y C, et al. Optical power control for GSNR optimization based on C+L-band digital twin systems [J]. *Journal of lightwave technology*, 2023, 42(1): 95–105. DOI: 10.1109/jlt.2023.3303783
- [2] CORREIA B, SADEGHI R, VIRGILLITO E, et al. Optical power control strategies for optimized C+L+S-bands network performance [C]//Proceedings of Optical Fiber Communication Conference (OFC) 2021. Optica Publishing Group, 2021: 1–3. DOI: 10.1364/ofc.2021.w1f.8
- [3] ROBERTS I, KAHN J M, HARLEY J, et al. Channel power optimization of WDM systems following Gaussian noise nonlinearity model in presence of stimulated Raman scattering [J]. *Journal of lightwave technology*, 2017, 35(23): 5237–5249. DOI: 10.1109/jlt.2017.2771719
- [4] SONG Y C, FAN Q R, LU C, et al. Efficient three-step amplifier configuration algorithm for dynamic C L-band links in presence of stimulated Raman scattering [J]. *Journal of lightwave technology*, 2023, 41(5): 1445–1453. DOI: 10.1109/jlt.2022.3223919
- [5] ZHANG C Y, WANG D S, JIA J W, et al. Potential failure cause identification for optical networks using deep learning with an attention mechanism [J]. *Journal of optical communications and networking*, 2022, 14(2): A122. DOI: 10.1364/jocn.438900
- [6] FAN Q R, ZHOU G, GUI T, et al. Advancing theoretical understanding and practical performance of signal processing for nonlinear optical communications through machine learning [J]. *Nature communications*, 2020, 11: 3694. DOI: 10.1038/s41467-020-17516-7
- [7] DELIGIANNIDIS S, BOGRIS A, MESARITAKIS C, et al. Compensation of fiber nonlinearities in digital coherent systems leveraging long short-term memory neural networks [J]. *Journal of lightwave technology*, 2020, 38(21): 5991–5999. DOI: 10.1109/jlt.2020.3007919
- [8] LIN X, LUO S H, SOMAN S K O, et al. Perturbation theory-aided learned digital back-propagation scheme for optical fiber nonlinearity compensation [J]. *Journal of lightwave technology*, 2022, 40(7): 1981–1988. DOI: 10.1109/jlt.2021.3133475
- [9] NIU Z K, YANG H, LI L, et al. Learnable digital signal processing: a new benchmark of linearity compensation for optical fiber communications [J]. *Light: science & applications*, 2024, 13: 188. DOI: 10.1038/s41377-024-01556-5
- [10] IP E, KAHN J M. Compensation of dispersion and nonlinear impairments using digital backpropagation [J]. *Journal of lightwave technology*, 2008, 26(20): 3416–3425. DOI: 10.1109/jlt.2008.927791
- [11] KARANOV B, CHAGNON M, AREF V, et al. Concept and experimental demonstration of optical IM/DD end-to-end system optimization using a generative model [C]//Proceedings of Optical Fiber Communication Conference (OFC) 2020. Optica Publishing Group, 2020. DOI: 10.1364/ofc.2020.th2a.48
- [12] NIU Z K, YANG H, ZHAO H C, et al. End-to-end deep learning for long-haul fiber transmission using differentiable surrogate channel [J]. *Journal of lightwave technology*, 2022, 40(9): 2807–2822. DOI: 10.1109/jlt.2022.3148270
- [13] LI L, NIU Z K, XIAO J Z, et al. Joint pre- and post-learned perturbation nonlinearity compensation optimization for long-haul optical fiber transmission based on end-to-end deep learning [C]//Proceedings of Optical Fiber Communication Conference (OFC) 2025. Optica Publishing Group, 2025. DOI: 10.1364/ofc.2025.w1k.6
- [14] LI Z Y, WANG C X, JIA J L, et al. Model-driven deep-learning for end-to-end optimization in fiber-terahertz communication systems [J]. *Journal of lightwave technology*, 2025, 43(7): 3099–3117. DOI: 10.1109/jlt.2024.3519360
- [15] AGRAWAL G P. Chapter 2 – pulse propagation in fibers [EB/OL]. [2025-11-10]. <https://www.oreilly.com/library/view/nonlinear-fiber-optics/9780128170434/B9780128170427000099.xhtml>
- [16] POGGIOOLINI P. The GN model of non-linear propagation in uncompensated coherent optical systems [J]. *Journal of lightwave technology*, 2012, 30(24): 3857–3879
- [17] POGGIOOLINI P, BOSCO G, CARENA A, et al. The GN-model of fiber non-linear propagation and its applications [J]. *Journal of lightwave technology*, 2014, 32(4): 694–721. DOI: 10.1109/jlt.2013.2295208
- [18] SERENA P, LASAGNI C, MUSSETTI S, et al. On numerical simulations of ultra-wideband long-haul optical communication systems [J]. *Journal of lightwave technology*, 2020, 38(5): 1019–1031. DOI: 10.1109/jlt.2019.2938580
- [19] WANG D S, SONG Y C, LI J, et al. Data-driven optical fiber channel modeling: a deep learning approach [J]. *Journal of lightwave technology*, 2020, 38(17): 4730–4743. DOI: 10.1109/jlt.2020.2993271
- [20] YANG H, NIU Z K, XIAO S L, et al. Fast and accurate optical fiber channel modeling using generative adversarial network [J]. *Journal of lightwave technology*, 2021, 39(5): 1322–1333. DOI: 10.1109/jlt.2020.3037905
- [21] HE X C, YAN L S, JIANG L, et al. Fourier neural operator for accurate optical fiber modeling with low complexity [J]. *Journal of lightwave technology*, 2023, 41(8): 2301–2311. DOI: 10.1109/jlt.2022.3229015

- [22] VASWANI A, SHAZER N, PARMAR N, et al. Attention is all you need [EB/OL]. [2025-11-02]. <https://www.tech-odyssey.cn/pdf/Al/Attention-is-all-your-need-ZH.pdf>
- [23] SHI M H, YANG H, NIU Z K, et al. Accurate and efficient optical fiber WDM transmission modeling using the encoder-only transformer with feature decoupling distributed method [C]// Proceedings of Asia Communications and Photonics Conference/ 2023 International Photonics and Optoelectronics Meetings (ACP/POEM). IEEE, 2023: 1-5. DOI: 10.1109/ACP/POEM59049.2023.10368835
- [24] ZHUANG B, LIU J, PAN Z, et al. A survey on efficient training of transformers [EB/OL]. [2025-11-10]. <https://arxiv.org/abs/2302.01107>
- [25] SU J L, AHMED M, LU Y, et al. RoFormer: enhanced transformer with rotary position embedding [J]. Neurocomputing, 2024, 568: 127063. DOI: 10.1016/j.neucom.2023.127063
- [26] BELTAGY I, PETERS M E, COHAN A. Longformer: the long-document transformer [EB/OL]. [2025-11-05]. https://milen.readthedocs.io/zh/latest/source/pretrain_model/longformer.html
- [27] MARCUSE D, MANYUK C R, WAI P K A. Application of the Manakov-PMD equation to studies of signal propagation in optical fibers with randomly varying birefringence [J]. Journal of lightwave technology, 1997, 15(9): 1735-1746. DOI: 10.1109/50.622902
- [28] EVANGELIDES S G, MOLLENAUER L F, GORDON J P, et al. Polarization multiplexing with solitons [J]. Journal of lightwave technology, 1992, 10(1): 28-35. DOI: 10.1109/50.108732
- [29] SHAO J, LIANG X J, KUMAR S. Comparison of split-step Fourier schemes for simulating fiber optic communication systems [J]. IEEE photonics journal, 2014, 6(4): 7200515
- [30] SHI M H, NIU Z K, YANG H, et al. Fast and accurate waveform modeling based on sequence-to-sequence framework for multi-channel and high-rate optical fiber transmission[J]. Optics Letters, 2025, 50(7): 2286-2289
- [31] SINKIN O V, HOLZLOHNER R, ZWECK J, et al. Optimization of the split-step Fourier method in modeling optical-fiber communications systems [J]. Journal of lightwave technology, 2003, 21(1): 61-68. DOI: 10.1109/JLT.2003.808628
- [32] YANG H, NIU Z K, ZHAO H C, et al. Fast and accurate waveform modeling of long-haul multi-channel optical fiber transmission using a hybrid model-data driven scheme [J]. Journal of lightwave technology, 2022, 40(14): 4571-4580. DOI: 10.1109/JLT.2022.3168698

作者简介



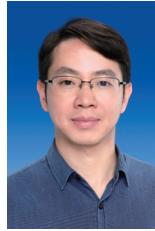
史明辉, 上海交通大学集成电路学院(信息与电子工程学院)在读博士研究生;研究方向为基于深度学习的光通信系统建模;发表论文5篇。



郑智雄, 上海交通大学集成电路学院(信息与电子工程学院)在读硕士研究生;主要研究方向为基于深度学习的光通信系统建模。



牛泽坤, 上海交通大学博士后;研究并开发的AI光纤信道模型实现上千倍的速度提升,合作成果已应用于华为光产品线,主导开发的中国首个智能光传输开源仿真平台(ifibertrans.sjtu.edu.cn),被近百家研究机构使用,中国电信评价其为“光通信仿真技术国产化的里程碑产品”;发表论文10余篇。



义理林, 上海交通大学教授、博士生导师,区域光纤通信网与新型光通信系统国家重点实验室副主任, OFC TPC成员(2017—2019),《Optical Fiber Technology》编委;主要从事智能光通信和智能光纤激光器方面的研究;所负责的智能锁模激光器工作被美国光学学会杂志《Optics & Photonics News》评为2019年度全球光学进展,并入选2019中国光学十大进展,2020年获国家杰出青年科学基金项目资助;发表论文100余篇。



面向智算中心互联的光算协同技术研究

Optical-Computing Collaborative Technology for Intelligent Computing Center Interconnection

谭艳霞/TAN Yanxia¹, 满祥锟/MAN Xiangkun¹,
吴绍辉/WU Shaohui², 张贺/ZHANG He¹, 徐博华/XU Bohua¹
(1. 中国联合网络通信有限公司研究院, 中国 北京 100048;
2. 北京市国防动员办公室, 中国 北京 100053)
(1. Research Institute of China United Network Communications Co., Ltd., Beijing 100048, China;
2. Beijing Municipal National Defense Mobilization Office, Beijing 100053, China)

DOI: 10.12142/ZTETJ.202506003
网络出版地址: <https://link.cnki.net/urlid/34.1228.TN.20251218.1429.002>
网络出版日期: 2025-12-19
收稿日期: 2025-10-15

摘要: 针对智算中心互联对光网络的新需求,结合当前智算网络发展现状,探讨智算中心互联架构及关键技术,以实现高性能算力互联。同时,针对跨智算中心分布式协同训练场景,搭建基于光传送网(OTN)的跨智算中心现网试验环境,在广域收敛比不低于16:1的场景下,百亿AI大模型跨域分布式训练性能达到95%以上。该试验验证采用单波800G实现300 km的传输,并验证其超高可靠传输能力。

关键词: 智算中心互联; 光传送网; 分布式协同训练; 高可靠传输

Abstract: The interconnection architecture and key technologies for intelligent computing centers are explored to address the new demands of optical networks for their interconnection, while considering the current development status of intelligent computing networks, with the aim of achieving high-performance computing power interconnection. Furthermore, focusing on the scenario of distributed collaborative training spanning multiple intelligent computing centers, an optical transport network (OTN)-based experimental testbed for cross-center interconnection is implemented on a live network. Under conditions where the wide-area convergence ratio is no less than 16:1, a performance of over 95% is achieved for cross-domain distributed training of AI large models with 10 billion parameters. Single-wave 800G transmission over 300 km is employed, and its ultra-high reliability and transmission capability are verified.

Keywords: interconnection of intelligent computing centers; optical transport network; distributed collaborative training; highly reliable transmission

引用格式: 谭艳霞, 满祥锟, 吴绍辉, 等. 面向智算中心互联的光算协同技术研究 [J]. 中兴通讯技术, 2025, 31(6): 13-19. DOI: 10.12142/ZTETJ.202506003

Citation: TAN Y X, MAN X K, WU S H, et al. Optical-computing collaborative technology for intelligent computing center interconnection [J]. ZTE technology journal, 2025, 31(6): 13-19. DOI: 10.12142/ZTETJ.202506003

随着人工智能(AI)大模型的快速发展,智算中心已成为支撑企业数字化转型的关键基础设施,也是推动AI大模型创新的核心动力。据统计,大模型参数规模每两年增长约10倍,算力需求持续攀升,而当前图形处理器(GPU)芯片算力仅以2~4倍的速度增长,远落后于模型规模的扩张速度。随着模型体量不断增大,单体智算中心在算力、电力、空间等方面面临限制,需要在园区乃至更广范围内整合多个智算中心资源,通过高速互联形成逻辑统一的超级算力资源池,以支持大规模分布式协同训练。

近年来,中国积极实施“东数西算”工程,加快建设全国一体化算力网络,有力推动了智算中心互联的技术演进与业务创新。通过光网络与算力资源的协同调度,实现泛在化算力聚合与一体化服务,形成“以网强算”的发展格局。在智算时代,光网络需满足跨中心数据推理、多数据中心协同训练等多样化任务需求,具备更弹性、更智能的网络能力。因此,智算中心互联对大带宽、低时延、高可靠及智能化的光网络提出了明确要求,也为光通信领域带来了新的发展机遇。

1 智算中心互联网络

为支撑大模型训练的持续增长需求, AI算卡规模正快速扩张。以OpenAI最新官宣启动的下一代前沿模型GPT-6为例, 其训练预计将需要高达70万~80万张算卡支持, 这已远超单体智算中心的承载能力。在此背景下, 分布式协同计算已成为必然趋势。

智算中心互联光网络正是为应对这一趋势而构建的, 旨在打造一个具备大带宽、低时延、高可靠、强智能特性的全光互联底座。该网络通过光算协同等关键技术, 实现对分布式算力资源的高效协同与智能调度, 能够有效满足“数据入算、模型训练、推理下发”等典型智算场景的需求, 并以此驱动新的业务增长。

1.1 智算中心光互联架构及典型场景

随着智能城市、自动驾驶和超高清视频直播等应用的快速发展, AI技术正加速向各行各业深度渗透, 区域间以及同城/区域内算力协同的需求急剧增长。智算中心互联架构如图1所示, 通过构建算力间全光高速平面, 将算力中心的Spine/Leaf节点经智算网关直接与光传送网(OTN)的光传输设备相连。智算网关具备多业务流识别与调度能力, 可提供长距离、高效率、智能化的流量调度功能; OTN设备则基于物理层参数与业务侧参数协同, 实现高吞吐、长距离、无损的数据传输服务, 从而保障智算节点间数据传输质量, 推动分布式算力资源的高效互联与协同发展。

从承载网络的视角分析, 结合智算中心互联架构, 可归

纳出四大典型需求场景, 如图2所示: 数据入算、模型训练、模型下发与模型推理^[9]。

1) 数据入算。数据入算指将各类数据上传至算力中心, 为后续训练与推理提供数据基础, 在医疗、政务、金融、影视、科研等领域均有广泛需求。在该场景下, 承载网络需具备弹性灵活的传输能力, 例如能够识别并高效调度大象流, 以应对太字节(TB)/拍字节(PB)级海量数据的传输需求。同时, 网络应支持根据业务实际需要动态调配资源, 提供弹性可定制、任务化的服务模式。

2) 模型训练。模型训练是指通过大量的数据和复杂的算法, 使模型学习数据中的规律和特征, 从而具备特定的智能决策能力或预测能力。目前主要存在3种训练场景: 数据中心(DC)内训练、跨DC协同训练和存算分离拉远训练。不同的训练场景对网络提出了差异化的要求。其中, 跨DC协同训练场景是基于多DC协同进行超大模型训练或实现碎片化算力整合出租, 对网络的运力服务提出了更大的挑战, 需要网络具备广域无损、低时延与负载均衡等能力; 存算拉远训练场景主要针对政务办公、医疗应用、金融领域等数据敏感用户, 实现用户数据私域存储与AIDC之间的高效拉远训练, 需要网络具备广域无损和数据安全保障等能力。

3) 模型下发。模型下发是指将训练完成的人工智能模型从开发环境部署至生产环境或各类终端设备, 以实现智能推理与决策能力。此场景对承载网络的需求与数据入算相似, 同样要求网络具备高效传输大规模数据的能力, 能够根据业务需求灵活调度资源, 并提供按需定制或任务式的数据

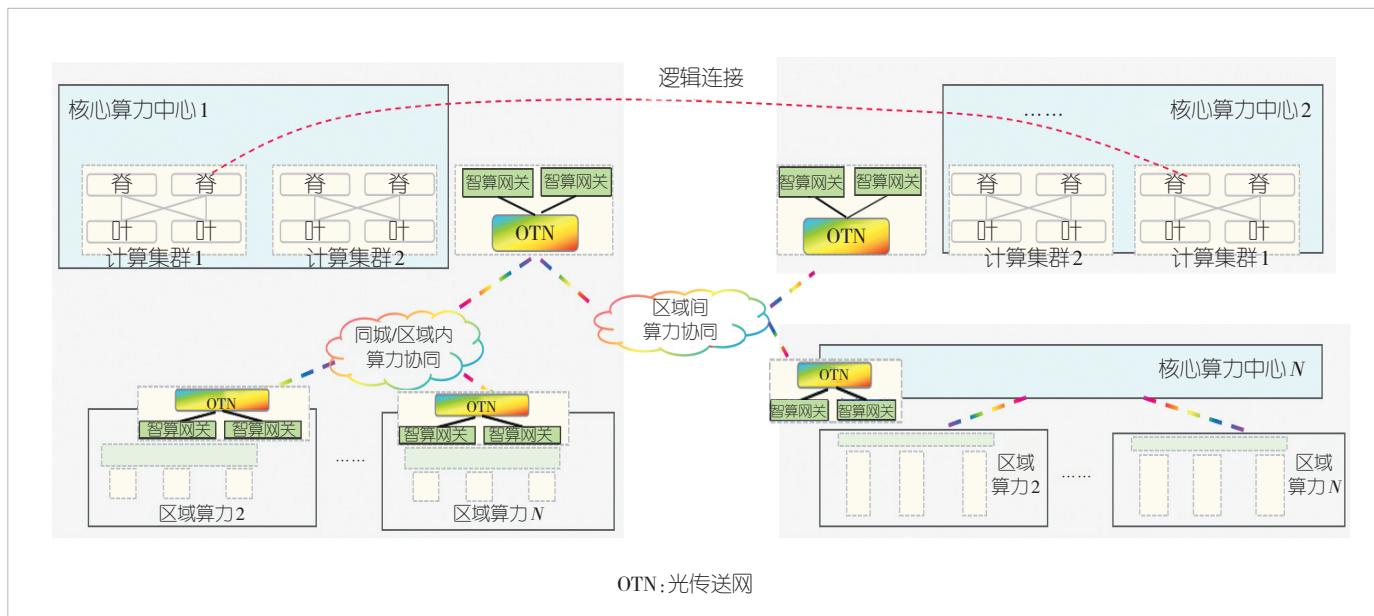
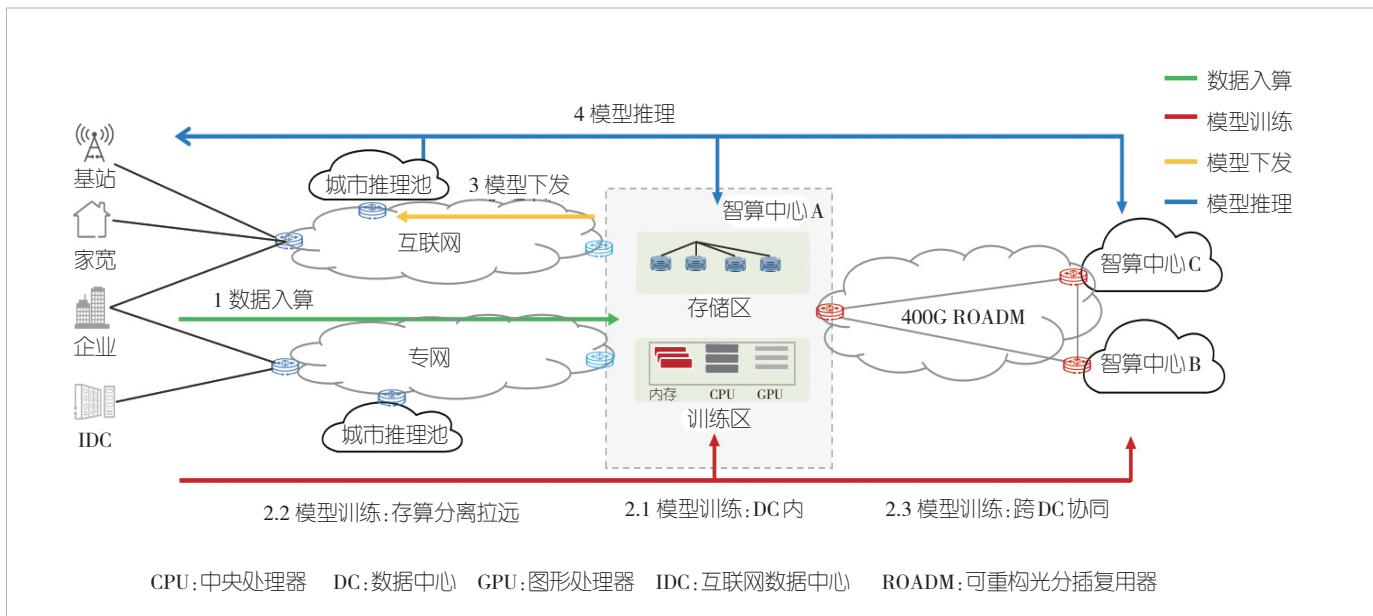


图1 智算中心互联架构示意图

图2 智算业务典型场景^[9]

传送服务模式。

4) 模型推理。模型推理是指将训练好的模型应用于实际业务场景, 对输入的新数据进行实时分析与预测, 从而提供决策支持或直接的服务输出。为满足用户泛在接入与实时交互的需求, 该场景要求承载网络具备广域覆盖能力及可定制的确定性承载性能。为此, 需借助应用感知、算力感知及算网一体化调度等关键技术, 实现网络服务的差异化与精准化承载。

1.2 智算中心光互联面临挑战

与智算中心内部互联相比, 智算中心之间的长距离互联环境更为复杂。为满足上述智算中心互联典型场景中的用户需求, 光网络首先需要具备长距离、大带宽、高速率的基础传输能力。在此基础上, 还需进一步提供弹性、按需、可靠、智能的连接服务, 并助力运营商实现新业务形态与商业模式的拓展。具体而言, 智算中心互联对光网络的关键需求主要包括^[8]:

1) 大带宽: 随着智算卡规模从千卡迈向万卡级别, 所需互联带宽常高达数百太比特至超拍比特量级, 单纤容量需持续提升。网络需兼顾大客户的超大带宽需求, 实现海量数据的快速迁移, 同时也要降低中小企业接入算力的成本, 提供灵活的带宽服务。

2) 低时延: 传输距离的增大会带来时延的累积。为满足跨智算中心协同训练对传输时延的严格要求, 光网络必须提供稳定且可保障的低时延传输能力。

3) 任务式灵活拆建: 为适应人工智能计算任务的动态需求, 需支持光网络连接的快速构建、灵活调整与及时拆除, 实现算力资源与网络资源的协同弹性调度, 从而满足业务灵活拆建的需求。

4) 可信可靠: 在跨智算中心互联场景下, 网络丢包、闪断及故障等问题可能影响协同计算的可行性和效率, 降低算力利用率, 甚至导致协同训练中断。与智算中心内部丢包相比, 长距离传输中的丢包重传机制可能引入高达千倍的时延累积^[10-11]。因此, 光网络需具备高可靠的无损传输能力, 并与终端侧协同优化, 保障高性能传输协议的效率。

5) 统一管理编排: 当前光网络与智算中心网络在管控上相互分离, 导致跨智算中心的分布式协同计算难以实现性能最优。现有固化的组织架构与生产流程亦无法支持逻辑统一的异地智算中心快速部署。因此, 需对现有管理编排体系、生产流程及管控平台进行改造与升级。

2 智算中心光互联关键技术

着眼于智算中心间互联需求, 实现算力节点之间的一体化协同、枢纽间/枢纽内算力的任意调度, 本文以智算中心间互联给光网络带来的挑战为基础, 提出智算中心光互联的关键技术。

2.1 确定性承载技术

智算中心间基于 OTN/可重构光分插复用器 (ROADM) 技术^[12-13], 采用单波 400G/800G 高速传输^[13]构建大容量底层

光网络，通过在智算中心之间建立光层直达通道，实现高速、海量数据的传输服务。该长距传输不经过传统2/3层交换机或路由器设备，而是通过光层一跳直达至对端智算中心出口交换机，从而在物理层面避免因队列调度机制引入的额外时延与丢包，构建高效无损的点对点传输通道。

2.2 安全可靠技术

为保证客户数据安全，智算中心间网络需采用更可靠的安全技术以提升客户体验。OTN技术基于光层（L0）的波长波分复用与电层（L1）的光通路数据单元（ODU）固定时隙对用户业务进行物理隔离，确保各项业务资源独享、互不干扰。同时，OTN支持传输加密，并可与量子密钥分发^[14]（QKD）等高安全性技术结合，实现保密通信。

OTN可提供多层次、全面的保护机制，具备高可靠保护能力，保障数据传输业务不中断。根据故障层次的不同，可分为电层保护、光层保护及光电协同保护^[10]。电层保护面向业务级，提供端到端1+1保护，倒换时间低于50 ms；对可靠性要求极高或光纤中断风险较大的场景，还可支持抗多次断纤的毫秒级保护能力。光层保护针对网络级光线路或节点故障，支持光传输段（OTS）1+1保护或光层自动交换光网络（ASON）保护，并具备确定性重路由能力，以降低同一故障引发的大规模业务中断风险。光电协同保护示意如图3所示，适用于对可靠性有更高要求的业务，通过在电层部署1+1保护、光层配置ASON动态恢复，实现协同防护，可抵御多次光纤故障，满足业务99.999%的可靠性要求。

2.3 任务式带宽技术

为适应智算业务的弹性带宽需求，传输管道需从静态分配模式向灵活拆建模式演进，从以年为单位的长期占用转变为支持小时级、天级的分时复用。为此，光网络应具备“任务式敏捷建链能力”与“弹性带宽调整能力”。

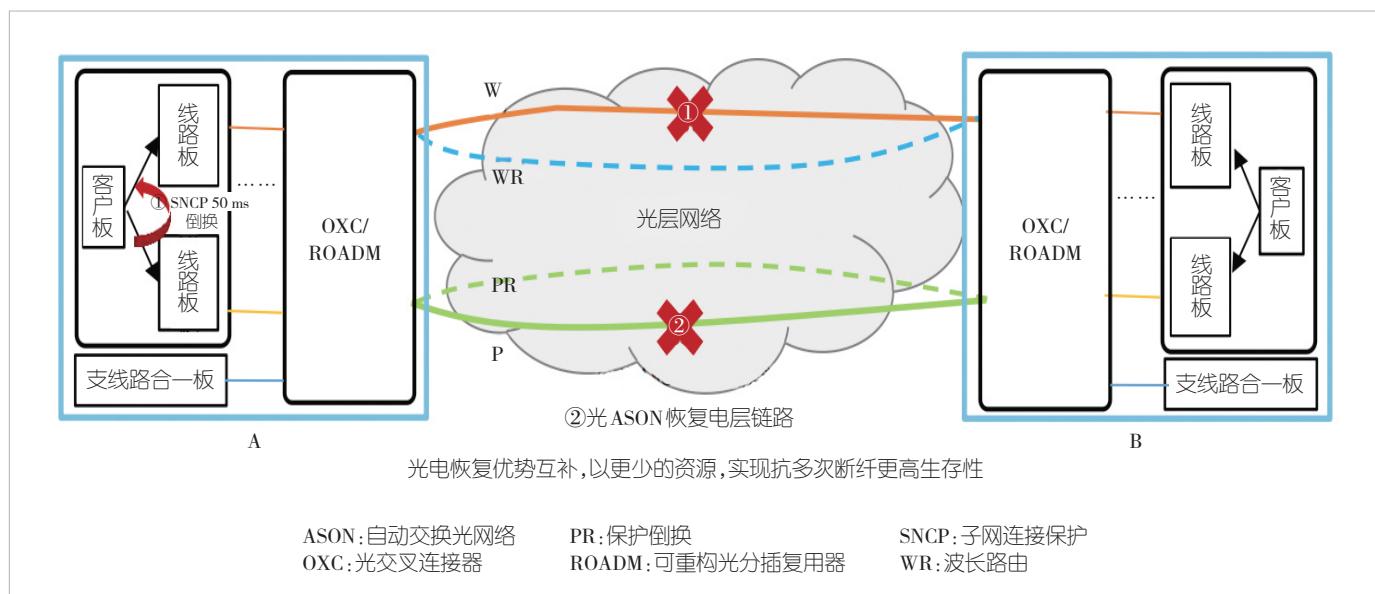
1) 任务式敏捷建链，实现波长级传输通道快速开通。任务式敏捷建链的核心在于构建两大关键能力：一是基于电层驱动、实现最优资源规划的光层控制能力；二是实现波长级业务从配置到开通的全流程自动化。

2) 基于Flex-O技术的弹性管道按需带宽调整。Flex-O技术通过信号拆分与FlexO映射，将多个标准速率的100G光模块绑定，实现 $N \times 100G$ 信号的高速传输。通过动态绑定或解绑定光模块数量，即可灵活调整传输带宽，满足业务对带宽的弹性需求。

2.4 长距离无损传输技术

为了应对超长距传输的挑战，承载网络需具备长距无损确定性传输能力，并与终端侧协同优化，以满足高性能协议对传输效率的要求。长距离无损传输技术主要包括端网协同物理层及保护路径信息、端网协同流控两方面。

1) 端网协同物理层及保护路径信息。在长距离场景下，可重构光分插复用器（ROADM）的吞吐量受传输距离、链路误码、保护倒换等因素影响。传输距离越长，吞吐量通常越低。网络可将物理路径的距离信息通过协议传递给网卡侧，网卡据此动态调整发送参数，以实现长距环境下的满速传输。当链路误码触发RDMA重传机制时，网络向终端侧通知



误码信息，终端据此判断重传原因属于链路误码还是网络拥塞，进而采取不同的传输策略。网络侧因故障触发的保护倒换可能影响传输效率，网络设备需将倒换后新路径的长度信息通知终端侧，由终端侧调整发送参数，以恢复满速传输。

2) 端网协同流控。在长距传输过程中，若远端数据中心发生拥塞，交换机会触发优先级流量控制(PFC)等流控机制，但其缓存能力通常不足以支撑长距无损传输。针对广域长距RDMA承载场景，OTN等传输设备需要具备对PFC流控信号的响应能力，协同缓存反压流量，并支持逐级向上游反压流量，从而确保RDMA流量在拥塞时不丢包，保障业务有效带宽。通过传输设备与交换机的协同机制，最终实现长距离环境下的无损传输^[4]。

2.5 算网协同管控编排技术

智算中心大规模部署后，呈现出分布式布局特征，且整体利用率较低。这一现状对算网协同智能调度提出了更高要求，需实现跨地域、跨层级、跨主体的高可靠调度。为保障典型智算业务的稳定运行，如图4所示，需通过算力管理系统与网络管理系统的协同调度，实现对算力与网络资源的统一管控与编排，从而为智算业务提供一体化的服务能力^[15]。

1) 构建算网业务运营层，实现算网产品的一体化运营服务。该层面向不同任务需求(如入算、训练、推理等)，为客户提供基于数据量、时长等参数的产品化服务新模式。

2) 构建算网协同编排调度层，实现算力与网络的统一编排。该层向上为运营层提供统一的算网一体业务服务能力，支持跨网络域与算力域的资源调配，还提供端到端业务编排与统一调度功能，支持时延优先、算力优先、成本优先等多种算网协同调度策略。

3) 强化网络管控层的动态拆建能力。网络管控系统需通过北向接口开放业务服务能力，将包含用户至各算力节点的连接时延、带宽等信息的“运力地图”同步至算网大脑，以支撑其快速、自动地选择最优算力节点与网络连接。

3 智算中心光互联的典型试验验证

针对多智算中心协同训练的典型应用场景，我们开展了跨智算中心的分布式协同训练现网试验验证。本次验证通过优化智算模型的并行策略，适配广域带宽条件下的超大收敛比，有效降低了超大规模智算中心互联场景中对长距离传输带宽的需求。试验中采用800G光传送网实现了300 km距离的超大带宽传输，并对多种可靠性影响因素进行了测试，包括智算拉远场景下传输带宽下降、链路误码以及保护倒换对计算效率的影响。

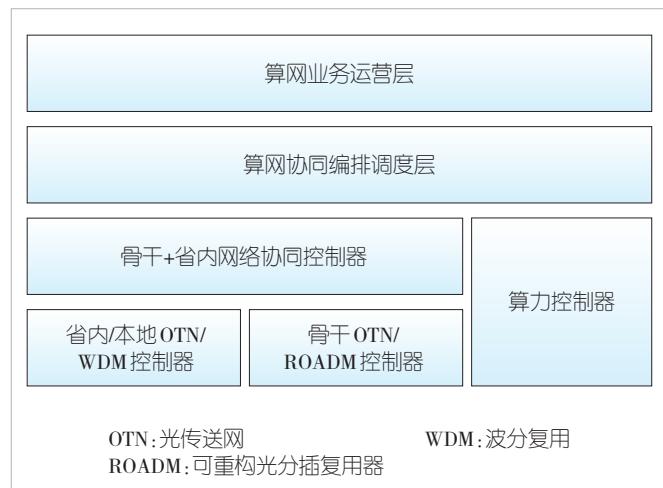


图4 算网协同管控编排架构

3.1 试验验证环境

分布式协同训练技术是指利用多个智算中心的计算资源协同完成同一AI大模型的训练任务，每个数据中心承担部分训练工作，并将每次迭代的结果实时同步，从而实现计算资源的按需分配与高效利用。跨智算中心分布式协同训练试验验证场景如图5所示，在现网验证中构建了包含32张计算卡的双智算中心环境，每个中心配置16张卡。智算中心内部采用Leaf+Spine两层拓扑的基于融合以太网的远程直接内存访问(RoCE)网络，中心之间通过智算网关经800G OTN设备直接互联，并依托OTN设备与长距光纤实现约329 km的远程连接，互联带宽为2×800G。试验验证中，OTN客户侧配置2个400G接口，在物理带宽上实现了4:1的收敛比。

为评估跨智算中心集群对模型训练效率的影响，在试验验证中部署了开源模型LLaMA 7B与LLaMA 13B，对在不同收敛比及模型参数配置下的计算效率变化情况进行研究与验证。在计算拉远协同训练场景的等效算力时，以相同参数配置下、基于数据中心内部网络进行的基准测试结果为参照，将拉远场景在同一时间内完成的训练进度占比作为等效算力指标。无论是否采用拉远部署，对于相同模型参数及训练样本集而言，其单轮迭代所需的总浮点运算次数保持不变，因此等效算力的计算公式可简化为：

$$\text{拉远等效算力} = \frac{\text{拉远场景总秒均算力}}{\text{基准场景总秒均算力}} =$$

$$\frac{\text{单迭代总浮点运算次数}/\text{拉远场景单次迭代时间}}{\text{单迭代总浮点运算次数}/\text{基准场单次迭代时间}} =$$

$$\frac{\text{基准场景单次迭代时间}}{\text{拉远场景单次迭代时间}} \%。$$

3.2 试验验证结论

本次试验通过引入新一代智算网关、精准流量控制与并行策略优化，在按张量并行 (TP)：流水线并行 (PP)：数据并行 (DP) = 4 : 4 : 2 的比例划分模型的情况下，采用 PP 并行方式跨中心拉远部署，实现了不低于 16 : 1 的广域收敛比。在此配置下，百亿参数规模大模型的分布式训练性能可达到单智算中心训练性能的 95% 以上。算力效率测试结果如图 6 所示，图中数据为多次测试取平均所得。

在光传送网可靠性对算效影响方面，通过在 OTN 网络中模拟传输带宽下降、链路误码及保护倒换等故障，研究并验证了不同故障场景对计算效率的影响。

在服务器上部署 LLaMA2 13B 模型，按 TP : PP : DP = 4 : 2 : 4 的比例划分模型，并行方式分别配置为 PP 拉远与 DP 拉远。在 PP 拉远场景下，当传输带宽从 800G 降至 400G 时，单轮迭代时间为 10.840 12 s，对算效基本无影响。在 DP 拉远场景下，同样将带宽从 800G 降至 400G，单轮迭代时间在一个迭代周期内增加至 15.249 3 s，计算效率下降约 38.496%，随后逐渐恢复。

我们将网关收敛比设置为 4 : 1，并行方式配置为 PP 拉远。无误码时单轮迭代时间正常约为 10.784 3 s；当出现误码且误码率在 10^{-9} 至 10^{-10} 量级间波动时，模型训练效率下降，单轮迭代时间延长至约 15.42 s，如图 7 所示。当误码率升至 10^{-8} 量级时，模型训练将出现中断。

我们将网关收敛比设置为 8 : 1、并行方式配置为 PP 拉远。当模拟工作链路发生 50 ms 级别保护倒换时，仅影响一个训练迭代周期，该周期内计算效率下降约 56%，但对整体协同训练过程影响有限。当模拟工作链路触发 ASON 动态重路由倒换时，同样影响一个迭代周期，但在部分情况下会导致模型训练直接中断。训练中断的出现，可能与倒换过程中关键数据帧丢失有关。

4 结束语

面向智算中心互联的光算协同技术，需以光网络为物理

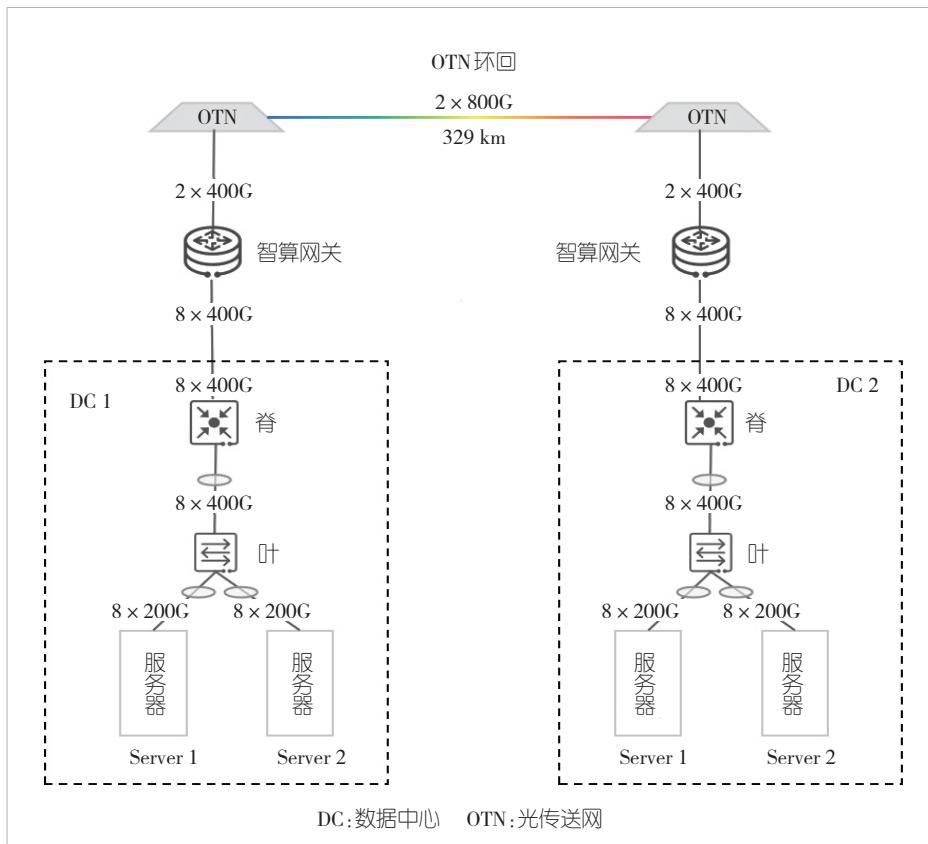


图 5 跨智算中心协同训练试验验证拓扑图

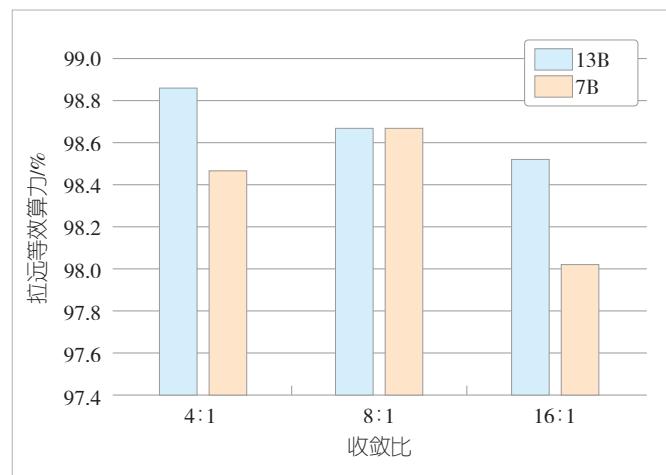


图 6 分布式训练等效算力测试结果

底座，通过突破确定性承载、安全可靠、任务式带宽、长距离无损传输以及算网协同管控编排等关键技术，构建“高速无损、灵活智能”的算力互联体系。未来，为应对智算网络面临的多重挑战，将在现有研究成果基础上，积极探索任务式灵活部署产品形态，优化算网协同管控机制与网络计费模式，实现算力在任意地点与时间的灵活调度。同时，持续开

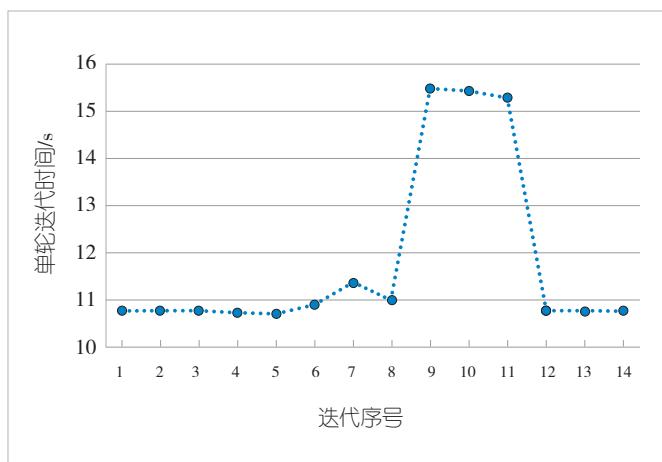


图7 链路误码对训练单轮迭代时间的影响测试结果

开展异构算力协同训练应用试验, 推动RDMA协议层与光网络物理层之间的感知联动, 最终实现超长距离环境下的高吞吐量无损传输。

参考文献

- [1] 李国杰. 智能计算技术的历史性突破与巨大挑战 [J]. 集成技术, 2025, 14(1): 1–8
- [2] 中国信息通信研究院. 中国算力发展报告(2024年) [R]. 2024
- [3] 丁宏庆, 张鹏飞, 牛红, 等. 云化的智算中心万卡集群创新与实践 [J]. 电信科学, 2024, 40(12): 125–135
- [4] TAN Y X, MAN X K, WANG G Q, et al. Field trial of long-distance RDMA lossless transmission for wide-area data center interconnection [EB/OL]. [2024-11-05] [2025-11-08]. <https://ieeexplore.ieee.org/document/10809882>
- [5] 张德朝, 孙将, 曹珊, 等. 面向跨智算集群互联的新型HIC-OTN技术 [J]. 电信科学, 2025, 41(4):53–60
- [6] LIU Y Y, ZHANG A X, WANG X S, et al. Field trial of multi-datacenter distributed training for LLM based on bandwidth convergence and two parallel strategies over 120km high-reliability 800Gbit/s C+L OTN [EB/OL]. [2025-11-09]. <https://ieeexplore.ieee.org/document/11047207>
- [7] 中国信息通信研究院. 算力时代全光网架构研究报告 [R]. 2024
- [8] 中国联通研究院. 基于RDMA的长距无损数据搬移技术白皮书 [R]. 2024
- [9] 易昕昕, 张乃晗, 刘雅承, 等. 算力智联网关键技术研究 [J]. 中兴通讯技术, 2025, 31(2): 31–38. DOI: 10.12142/ZTETJ.202502005
- [10] 王光全, 满祥锟, 徐博华, 等. 确定性光传输支撑广域长距算力互联 [J]. 邮电设计技术, 2024(2): 7–13
- [11] MACARTHUR P, RUSSELL R D. A performance study to guide RDMA programming decisions [EB/OL]. [2025-11-09]. <https://ieeexplore.ieee.org/document/6332248>
- [12] 唐雄燕, 王海军, 杨宏博. 面向专线业务的光传送网(OTN)关键技术及应用 [J]. 电信科学, 2020, 36(7): 18–25
- [13] WANG C Y, HU Y K, SHEN S K, et al. Channel power management of 400 G transmission system based on C6T + L6T spectrum and QPSK modulation format [J]. Optics express, 2024, 32(11): 20279. DOI: 10.1364/oe.523644
- [14] QU W X, ZHANG Y, LU Y M, et al. Low-cost lightweight-client twin-field quantum key distribution network with wavelength division multiplexing [EB/OL]. [2025-11-10]. <https://ui.adsabs.harvard.edu/abs/2022OptEn..61a6102Q/abstract>
- [15] 中国联通研究院. AI时代的全光底座白皮书 [R]. 2025

作者简介



谭艳霞, 中国联合网络通信有限公司研究院高级工程师; 主要研究方向为传输网管控技术、智算网络等。



满祥锟, 中国联合网络通信有限公司研究院正高级工程师; 主要研究方向为光通信网络、智算网络等。



吴绍辉, 北京市国防动员办公室工程师; 主要研究方向为传输网管控技术、智算网络等。



张贺, 中国联合网络通信有限公司研究院正高级工程师; 主要研究方向为光纤传输、同步网新技术等。



徐博华, 中国联合网络通信有限公司研究院高级工程师; 主要研究方向为数据中心网络关键技术等。



50G-PON 平滑演进的挑战和方案研究

Challenges and Solutions for Smooth Evolution of 50G-PON

杨波/YANG Bo¹, 张德智/ZHANG Dezhⁱ², 刘波/LIU Bo¹

(1. 中兴通讯股份有限公司, 中国 深圳 518057;

2. 中国电信研究院, 中国 北京 102209)

(1. ZTE Corporation, Shenzhen 518057, China;

2. China Telecom Research Institute, Beijing 102209, China)

DOI: 10.12142/ZTETJ.202506004

网络出版地址: <https://link.cnki.net/urlid/34.1228.TN.20251230.1453.003>

网络出版日期: 2025-12-30

收稿日期: 2025-11-01

摘要: 提出了 50G-PON 三代时分共存系统架构, 以解决以太网无源光网络 (EPON)、10G-EPON 向 50G-PON 演进中因兼容现网非收窄 EPON 终端所面临的波长冲突等关键问题, 实现了 EPON 收窄/非收窄、10G-EPON 对称/非对称以及 50G-PON 对称/非对称 6 类终端全兼容共存。研究涵盖下行 $1.366 \pm 2 \text{ nm}$ 新波长与上行 1.25 Gbit/s、10.312 5 Gbit/s、24.883 2 Gbit/s、49.766 4 Gbit/s 四速率时分接收等物理层关键技术, 并完成了相应光模块与系统设备的研制及验证测试。结果表明, 所研光模块与系统指标满足现网 29 dB 功率预算要求并留有余量, 可支持现网光分配网络 (ODN) “零改动” 平滑升级。本研究成果为万兆光网演进提供了高效益、低成本且可持续的解决方案。

关键词: 50G-PON; EPON; 10G-EPON; 时分复用; 三代共存; 四速率突发接收机

Abstract: A three-generation time-division coexistence system architecture for 50G-passive optical network (PON) is proposed to address challenges including wavelength conflicts during the evolution from Ethernet PON (EPON) and 10G-EPON to 50G-PON, while ensuring backward compatibility with deployed non-narrowed EPON terminals. This architecture enables full coexistence among six classes of terminals: narrowed and non-narrowed EPON, symmetric and asymmetric 10G-EPON, and symmetric and asymmetric 50G-PON. Key physical-layer technologies were investigated, such as a newly defined $1.366 \pm 2 \text{ nm}$ downstream wavelength and an upstream four-rate (1.25 Gbit/s, 10.312 5 Gbit/s, 24.883 2 Gbit/s, and 49.766 4 Gbit/s) time-division reception scheme. Corresponding optical modules and system equipment for the three-generation coexistence were developed and validated. Test results show that the optical interface specifications of the module and the system performance satisfy the existing network's 29 dB power budget requirement with sufficient margin, thereby supporting a smooth, zero-touch evolution of the installed optical distribution network (ODN). An efficient, cost-effective, and sustainable solution is thus provided for the evolution toward 10G-class optical access networks.

Keywords: 50G-PON; EPON; 10G-EPON; time-division multiplexing; three-generation coexistence; four-rate burst receiver

引用格式: 杨波, 张德智, 刘波. 50G-PON 平滑演进的挑战和方案研究 [J]. 中兴通讯技术, 2025, 31(6): 20-28. DOI: 10.12142/ZTETJ.202506004

Citation: YANG B, ZHANG D Z, LIU B. Challenges and solutions for smooth evolution of 50G-PON [J]. ZTE technology journal, 2025, 31(6): 20-28. DOI: 10.12142/ZTETJ.202506004

随着云计算、4K/8K 视频、工业互联网等新兴业务的爆发式增长, 用户对网络带宽的需求呈现指数级攀升趋势。传统光接入网络已难以满足未来 10 年“万物智联”场景的承载需求。无源光网络 (PON) 作为光纤接入网主流技术, 经历从吉比特无源光网络 (GPON) 向 10 Gbit/s 无源光网络 (XG-PON) 和 10 Gbit/s 对称无源光网络 (XGS-PON) 的演进, 以及从以太网无源光网络 (EPON) 向 10 Gbit/s 以

太网无源光网络 (10G-EPON) 的升级, 并正逐步向更高速率的下一代 PON 迈进。50 Gbit/s 无源光网络 (50G-PON) 作为下一代无源光网络的核心技术, 凭借其单波长 50 Gbit/s 的超高带宽、低时延和高可靠性等优势, 已成为下一代 PON 的主流演进方向^[1-2]。

在 PON 技术的发展过程中, 国际电信联盟 (ITU) 主导的 GPON 与电气电子工程师协会 (IEEE) 主导的 EPON 长期并行发展, 形成了两大技术体系。两者在技术标准、设备生态和运营模式上存在较大差异, 增加了网络升级与运维的复杂性。2018 年, ITU 电信标准化部门 (ITU-T) 正式立项启

基金项目: 上海科委启明星项目 (23QB1405800)

动 50G-PON 标准的研制工作，将其确立为 XG-PON 和 XGS-PON 的下一代演进方向，相关标准包括：G.9804.1^[3]、G.9804.2（通用传输汇聚层）^[4]以及 G.9804.3（单波长 50G-PON 物理层^[5]）。2019—2025 年，支持单波 50G 的 50G-PON 系列标准及其修正案陆续完成研究与发布^[6]。与此同时，IEEE 于 2020 年完成了 $N \times 25G$ -EPON 标准的制定^[6]，该标准单波长支持 25 Gbit/s 速率，可通过双波长绑定实现 50 Gbit/s 传输。 $N \times 25G$ -EPON 技术主要受到北美运营商的关注。为凝聚产业链力量、共同推动关键技术的成熟，将 ITU-T GPON 体系与 IEEE EPON 体系归一化融合至 50G-PON 的路线，已获得全球范围内多数运营商的认可。因此，50G-PON 已成为 XG-PON、XGS-PON 及 10G-EPON 网络共同的下一代演进目标。

ITU GPON 体系在向 50G PON 演进过程中，其共存方案涉及 3 个上行波长选项：1 260~1 280 nm（选项 1）、1 290~1 330 nm（选项 2）以及 1 284~1 288 nm（选项 3）^[5]。其中，选项 1 仅支持 GPON 与 50G-PON 两代共存，选项 2 仅支持 10G-PON 与 50G-PON 两代共存，而选项 3 可支持 GPON、10G-PON 与 50G-PON 三代共存^[5]。因此，目前业界在 50G-PON 上行波长选择上普遍都采用适用场景更广泛的选项 3。

相比之下，IEEE EPON 体系的上行波长分布较为复杂。10G-EPON 对称系统上行波长为 1 260~1 280 nm；而对于非对称系统，EPON 与 10G-EPON 的上行波长存在收窄与非收窄两种类型。其中，EPON 非收窄类型的工作波长范围较宽，为 1 260~1 360 nm^[7]；收窄类型则为 1 290~1 330 nm。10G-EPON 非对称系统的上行波长在规模部署初期已收窄至 1 260~1 280 nm^[8]。

若 EPON、10G-EPON 与 50G-PON 采用现有 ITU-T 无源光网络共存标准 G.9805^[9] 所定义的波分复用架构，则会出现以下波长冲突问题：

1) 50G-PON 上行波长 $1 286 \pm 2$ nm 与 EPON 非收窄终端上行波长 1 260~1 360 nm 部分重叠，导致波分复用器件无法通过波长区分两种制式的上行信号，从而造成业务干扰。

2) 50G-PON 下行波长 $1 342 \pm 2$ nm 同样与 EPON 非收窄终端上行波长 1 260~1 360 nm 存在重叠，50G-PON 下行光信号经光分配网络（ODN）反射后可能进入 EPON 上行接收机，影响 EPON 上行业务。

3) 50G-PON 光网络单元接收机在 1 330~1 356 nm 范围内未定义明确的滤波要求^[5]，该波长范围内的 EPON 非收窄终端宽谱上行信号经 ODN 部分反射后可能进入 50G-PON 下行接收端，影响其接收灵敏度。

根据中国电信的调研，目前现网在用的 EPON 终端总量超过 6 000 万台。按照网络演进趋势预测，至 2030 年，预计仍有超过 2 000 万台终端在网运行，短期内难以实现 EPON 终端的全面退网。日本、韩国等国家的运营商也存在类似情况。

综合考虑标准定义与运营商光接入网现状，EPON、10G-EPON 与 50G-PON 三代系统若采用波分复用方式共存，将面临上下行波长冲突以及大量在网非收窄型 EPON 终端替换成本高等问题。

本文聚焦于 EPON 体系向 50G-PON 平滑演进的核心问题，提出一种 50G-PON 三代时分共存方案及相应系统架构，开展其物理层关键技术研究，并完成了基于 50G-PON 三代时分复用 Combo（组合）光模块的系统验证。研究结果表明，该时分共存方案可兼容 EPON、10G-EPON 及 50G-PON 三代制式，并支持非收窄 EPON、收窄 EPON、非对称 10G-EPON、对称 10G-EPON、非对称 50G-PON 和对称 50G-PON 共 6 类终端同时共存。所使用的 Combo 光模块光收发指标满足现网 29 dB 功率预算要求，从而为 EPON 及 10G-EPON 网络向 50G-PON 的平滑演进提供了理论依据。

1 50G-PON 三代时分共存系统架构

1.1 系统架构

50G-PON 三代时分共存系统架构如图 1 所示，其与波分共存方案的核心区别在于上行方向采用时分复用模式，分时隙接收不同类型终端发送的上行信号以避免上行波长冲突，彻底解决万兆光网无法兼容现网大量未收窄 EPON 终端的这一难题，实现了 EPON 终端全兼容共存，现网终端无须退网或新建 ODN。

50G-PON 三代时分共存系统采用多 PON 模块（MPM）共存模式，通过光线路终端（OLT）内置 50G-PON 三代时分复用 Combo 光模块以及 OLT 多媒体接入控制（MAC）来实现统一时分调度动态带宽分配（DBA）机制，以支持 EPON、10G-EPON 和 50G-PON 三代共存，非收窄 EPON、收窄 EPON、非对称 10G-EPON、对称 10G-EPON、非对称 50G-PON 和对称 50G-PON 6 类终端全兼容。其中，EPON 下行单波长支持 1.25 Gbit/s，上行单波长支持 1.25 Gbit/s；10G-EPON 下行单波长支持 10.3125 Gbit/s，上行单波长支持 1.25 Gbit/s、10.3125 Gbit/s 两种速率；50G-PON 下行单波长支持 49.7664 Gbit/s，上行单波长支持 24.8832 Gbit/s 和 49.7664 Gbit/s 两种速率；三代 PON 上下行均采用不归零（NRZ）线路编码。上行方向，50G-PON 三代时分共存系统

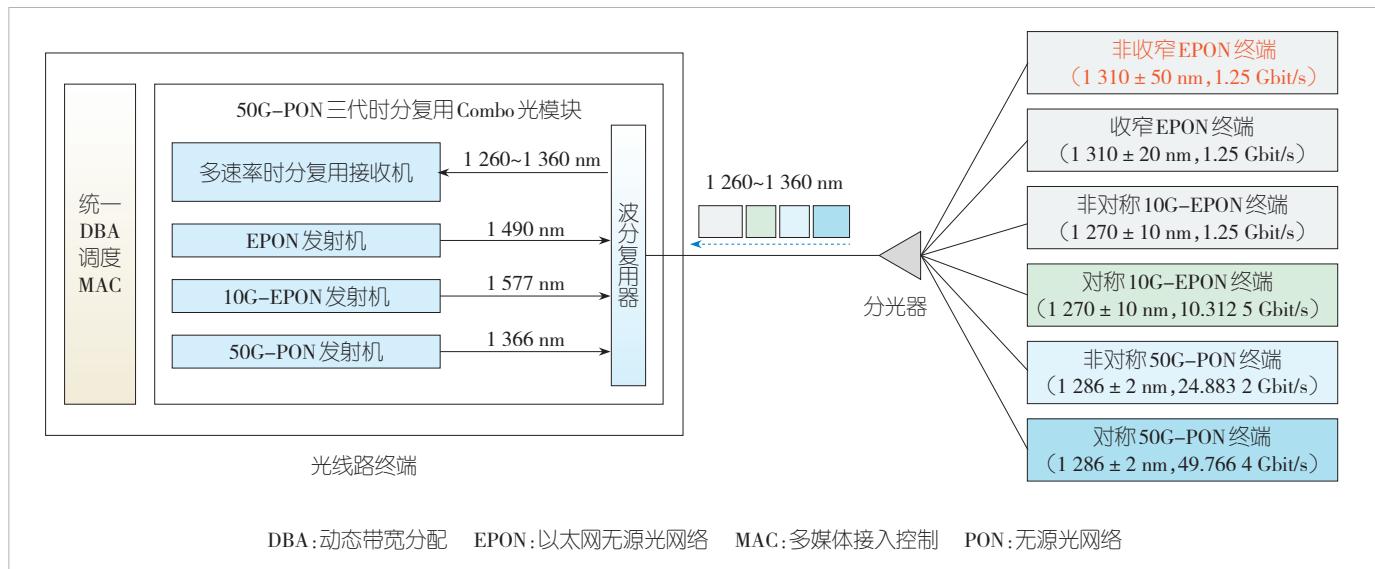


图1 50G-PON三代时分共存系统架构

则采用了四速率时分复用接收机制，接收机的工作波长范围为1 260~1 360 nm，分时隙接收1.25 Gbit/s、10.312 5 Gbit/s、24.883 2 Gbit/s和49.766 4 Gbit/s 4类速率上行光信号，解决EPON非收窄终端与50G-PON上行波长冲突的问题。下行方向，50G-PON三代时分共存系统采用传统波分复用架构，EPON、10G-EPON和50G-PON下行业务数据分别通过不同波长光信号传输至光网络单元(ONU)，并通过不同PON制式ONU内置的下行波长滤波器选择接收。为避免ODN反射引起的50G-PON下行波长1 342 ± 2 nm与EPON和10G-EPON非对称上行波长冲突问题，三代时分共存方案中50G-PON下行采用了新的波长定义。采用上述50G-PON三代时分共存的方案和系统架构，各制式下行带宽与波分方案一致，上行带宽通过统一DBA调度，按需分配给6类终端，带宽利用率更高。

1.2 波长规划和光接口指标定义

针对EPON、10G-EPON向50G-PON演进过程中遇到的各类波长冲突问题，50G-PON三代时分共存系统架构上行方向采用四速率时分接收方式，解决了50G-PON上行波长与EPON上行波长冲突问题。新定义的50G-PON下行波长，要求其在与EPON、10G-EPON和50G-PON上行波长，以及EPON、10G-EPON下行波长互不冲突的同时，满足50G-PON系统下行传输功率预算和传输距离要求。ITU-T G.9804.3标准规定50G-PON下行49.766 4 Gbit/s信号经过20 km光纤传输后，下行光路径代价(OPP)最大不超过3.5 dB，发射机色散和眼图闭合(TDEC)最大不超过5 dB。

对于50G-PON系统，其OPP主要来源于光纤传输引入的色散代价，另外还包括一些特定场景下非线性效应代价和0.5 dB规模应用余量。传统EPON和10G-EPON下行波长所处的S、C和L波段由于光纤色散系数较大，已不能支持49.766 4 Gbit/s速率20 km传输。在50G-PON三代时分共存的系统中，除去上行波长占用的O波段外，50G-PON下行波长需要选择更为靠近0色散波长的E波段。传统的E波段(1 360~1 460 nm)由于水峰的存在，通常不被用于高速光通信。但接入侧光纤通常部署年份较晚，光纤工艺的改进使得光纤在1 360~1 460 nm范围内的衰减显著降低^[10]，这扩展了光纤在E波段的应用。如ITU-T G.694.2中定义了粗波分复用(CWDM)18个标准波长通道，1 371 nm、1 391 nm、1 411 nm、1 431 nm和1 451 nm等波长位于E波段^[11]。其中，1 371 nm通道已在移动前传网络中广泛应用。

在E波段中，为规避水峰衰减峰值(1 383 nm)及其邻波长区域，同时降低下行色散代价，EPON、10G-EPON与50G-PON三代共存系统中的50G-PON下行波长应尽可能选取靠近O波段的波长。如图2所示，本文所述系统将50G-PON下行波长定义为1 366 ± 2 nm。其中，1 360~1 364 nm之间保留了4 nm的波长间隔，用作3类PON制式上下行波长在合分波器中的过渡带，以实现主干光纤中上下行波长的单纤双向传输，并有效隔离下行50G-PON信号的反射波长进入上行时分接收机，从而解决50G-PON下行波长与EPON非收窄上行波长的冲突问题。需要指出的是，若发射端采用常规50G电吸收调制激光器(EML)，在1 366 ± 2 nm波长下，其色散代价可能超过50G-PON标准规定的3.5 dB OPP

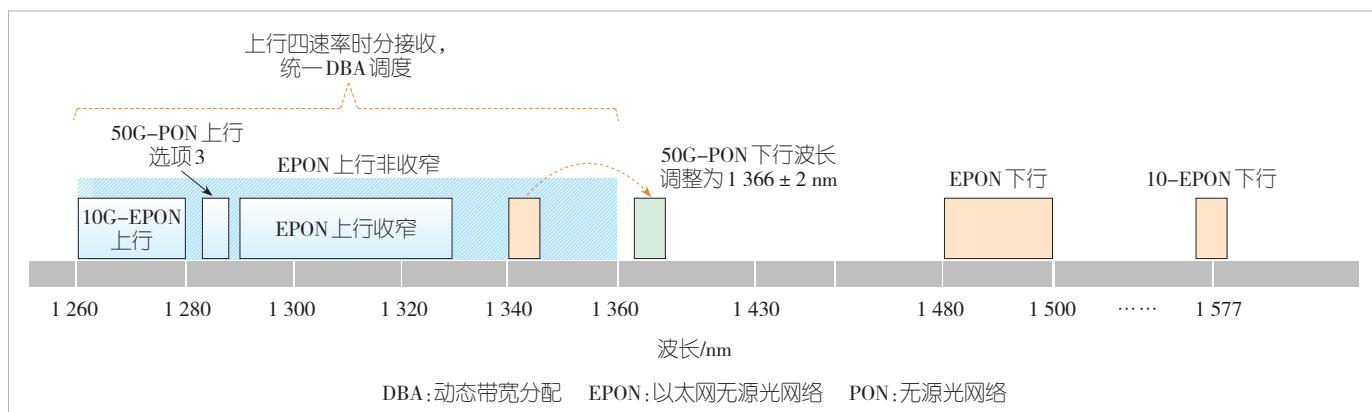


图2 50G-PON三代时分共存系统波长规划

限值。为降低该新下行波长的传输代价,可在下行50G-PON ONU接收侧采用增强型数字信号处理算法,或在发射侧采用负啁啾EML激光器方案^[11]。本研究将采用后一方案,即使用负啁啾1366 nm EML激光器。

在G.9804.3标准中,50G-PON Class N1等级规定的ODN功率衰减范围为14~29 dB,与EPON体系中的PX30、PR30及PRX30等级(15~29 dB)的预算范围最为接近。对于50G-PON三代时分共存系统的MPM光模块的关键光接口指标,除下行波长调整为 1366 ± 2 nm外,50G-PON的最小发射光功率、光调制幅度(OMA)及TDEC等其他光接口参数均参照G.9804.3标准Class N1等级的指标。EPON与10G-EAPON(对称/非对称)则分别遵循IEEE 802.3标准中PX30、PR30及PRX30等级的相应规定。为满足上述光接口指标要求,除需依托50G-PON已有的一系列关键技术外,如高功率发射机、高灵敏度接收机、高性能数字信号处理(DSP)均衡及前向纠错(FEC)等外,对于时分三模Combo光模块而言,其物理层关键技术还包括下行1366 nm发射机的色散代价控制以及上行四速率接收机的实现。

1.3 统一DBA调度

DBA是PON网络中实现上行带宽动态分配的核心机制,旨在提升带宽利用率、降低传输时延并同时更好地保障业务服务质量。在EPON、10G-EAPON与50G-PON三代共存的网络中,由于IEEE EPON体系与ITU-T GPON体系在突发机制、时钟域及DBA管理机制上存在差异,为实现不同体系PON制式的协同工作,必须采用统一的DBA机制进行整体调度。

实现不同PON制式的统一DBA调度,主要有两种可行机制。一种是将上行时隙按GPON和EPON两种制式划分为两个区域,在各区域内沿用其原有的DBA调度机制。该方案的实现较为简单,但难以最优地平衡带宽利用率与低时延

特性。另一种是将不同PON模式(如EPON、10G-EAPON、50G-PON等)的业务优先级统一映射到系统内部定义的调度优先级,从而在共存网络中实现带宽的公平调度,并可精细优化系统带宽利用率与低时延性能。本文中,我们提出一种具体的GPON DBA向EPON DBA的映射机制,GPON传输容器(T-CONT)类型1映射为EPON固定带宽,T-CONT类型2映射为EPON保证带宽,T-CONT类型3保证带宽和非保证带宽分别映射为EPON保证带宽和尽力而为带宽优先级0,T-CONT类型4映射为EPON尽力而为带宽优先级非0,T-CONT类型5的固定和保证带宽分别映射为EPON的固定带宽和保证带宽,T-CONT类型5的其他类型带宽映射为EPON尽力而为带宽,并选择优先级非0。通过该映射机制,可将GPON的T-CONT调度纳入EPON的统一调度框架,利用统一的DBA调度器同时处理EPON的逻辑链路标识(LLID)与GPON的T-CONT,并将GPON的T-CONT类型与EPON的优先级进行对应映射,从而实现GPON与EPON体系在共存网络中的统一DBA机制。如图1所示,在50G-PON三代时分共存系统中,OLT内部的三代时分共存MAC芯片集成了统一的DBA调度器。该调度器接收来自GPON(经适配后)和EPON的带宽请求,并基于统一的优先级与带宽策略进行调度,进而生成统一的上行带宽授权消息下发至各ONU,从而实现灵活且集中控制的带宽授权。借助该统一DBA调度机制,系统可支持EPON、10G-EAPON与50G-PON三代6类终端的上行时分复用,彻底解决了三代PON共存时上行波长冲突的问题。统一的DBA机制不仅简化了带宽管理与调度逻辑,避免了不同PON体系间的资源冲突,还能够按需动态分配带宽,从而提升整体网络效率。通过下行波分复用、上行时分复用以及协议机制的统一,50G-PON三代时分共存系统为运营商提供了平滑、灵活的网络演进路径。

2 50G-PON三代时分共存物理层关键技术

2.1 下行 1 366 nm EML 激光器传输特性

在 50G-PON 三代时分共存系统中, 将 50G-PON 下行工作波长范围调整为 1 364 ~ 1 368 nm 后, 为最大限度保持与 GPON 体系指标的兼容性, 除工作波长外, 其下行发射机性能仍需满足 G.9804.3 标准的相关要求。其中, 与波长密切相关的下行发射机关键指标包括 TDEC 与 OPP。TDEC 指标通常受到激光器 3 dB 带宽、调制系数、光信噪比 (OSNR)、啁啾系数以及非线性失真等一系列因素的影响。对于常规工作于 1 342 nm 的 50G-PON EML 激光器, 其 3 dB 带宽一般为 37.5 GHz, 消光比 (ER) 大于 7 dB (在带宽不受限时对应的调制系数约为 0.704)。当该激光器的啁啾系数为 0.5 时, 其 TDEC 及经 20 km 光纤传输后的色散代价分别约为 4.5 dB 与 2.5 dB^[1]。随着波长向长波长方向移动, G.652 光纤的色散系数逐渐增大, 20 km 传输对应的累积色散量从 1 342 ± 2 nm 时的 77.1 ps/nm 增加至 1 366 ± 2 nm 时的 116.1 ps/nm。若保持啁啾系数为 0.5, 1 366 nm EML 激光器的 TDEC 将超出 50G-PON 标准规定的 5 dB 上限。本文通过仿真来确定满足 TDEC < 5 dB 要求的 1 366 nm EML 激光器的啁啾系数取值范围。下行仿真链路如图 3 (a) 所示, 主要包括短应力伪随机码 (SSPR) 信号发生器、EML 激光器、20 km 光纤及 TDEC 计算模块。根据 G.9804.3 标准定义的测试方法, TDEC 计算模块中还包含光电转换器、滤波器及 TDEC 算法模块等^[5]。仿真中, 发射机的眼图直方图数据波形由 TDEC 计算模块中的光电探测器获取, 并通过滤波器将带宽限制在参考接收机定义的 18.75 GHz。TDEC 算法模块采用 13 阶前馈均衡器 (FFE) 作为虚拟参考均衡器, 对获取并滤波后的波形进行均衡, 最终根据均衡后的测试眼图闭合度以及均衡器噪声增强因子 C_{eq} 计算出 TDEC 值。

采用图 3 (a) 所示仿真链路及参数, 对不同啁啾系数在不同传输距离下的 TDEC 进行仿真, 结果如图 4 (a) 所示。仿真结果表明, 眇啾系数越小, 激光器的 TDEC 值越低。当啁啾系数小于或等于 0 时, 1 366 ±

2 nm EML 激光器可满足 50G-PON 标准中规定的、经 20 km 光纤传输后 TDEC ≤ 5 dB 的要求。

考虑到实际系统中抖动、串扰及非线性失真等因素会导致发射机眼图质量劣化, 并兼顾规模应用所需的余量, 建议将 1 366 ± 2 nm EML 激光器的啁啾系数设计在 -0.5~0。

在 50G-PON 系统中, 下行传输的另一关键参数为 OPP, 其主要由激光器经光纤传输引起的色散代价构成, 同时也包含少量多波长传输非线性效应代价与多路径反射代价。采用图 3 (a) 所示仿真系统与参数, 发射数据采用 15 阶伪随机二进制序列 (PRBS15), 接收侧 DSP 配置为 3 抽头判决反馈均衡 (DFE) 加 15 抽头 FFE 均衡算法, 对啁啾系数为 0 与 -0.2 的 1 366 ± 2 nm EML 激光器进行仿真, 其背靠背 (BtB) 与 20 km 传输后的结果如图 4 (b) 所示。仿真结果表明, 眇啾系数为 0 与 -0.2 的 1 366 ± 2 nm EML 激光器在 20 km 传输后的色散代价分别为 1.78 dB 与 1.08 dB, 低于啁啾系数为 0.5 的 1 342 nm EML 激光器在同等传输距离下的代价, 因而能够满足 50G-PON 时分三模系统下行 20 km 的传输要求。

通过对 1 366 ± 2 nm EML 激光器的 TDEC 与色散代价进行仿真分析可知, 当啁啾系数取值在 -0.5~0 时, 该激光器的下行发射 TDEC 指标以及 20 km 传输后的 OPP 均能满足 50G-PON G.9804.3 标准的要求。其发射光功率仅需满足标准中 OMA-TDEC ≥ 4.75 dBm 的规定即可。因此, 对于 1 366 ± 2 nm EML 激光器, 除需控制啁啾系数外, 无须对其发射光功

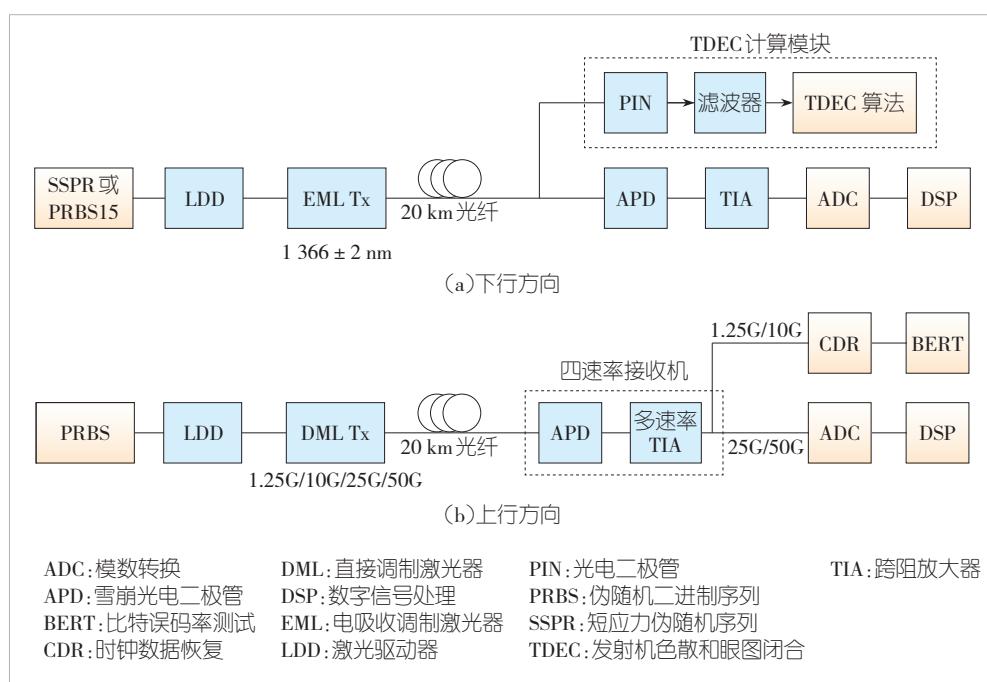
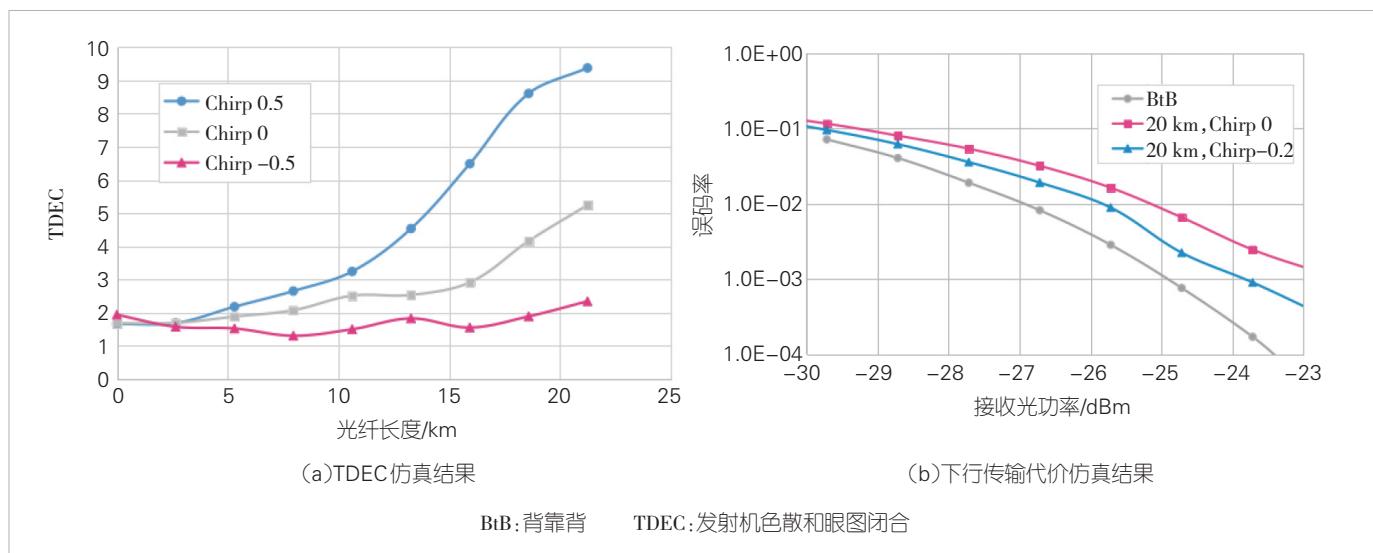


图 3 50G-PON 仿真链路

图4 不同啁啾系数 1366 ± 2 nm EML激光器

率、调制系数、OSNR等其他指标进行特殊设计。根据已有文献报道,通过适当调整电吸收调制器的吸收边与信号波长之间的相对关系,或优化EA偏置电压,可使高速EML激光器实现负啁啾特性,其值甚至可达-1以下^[12]。这表明,1366 \pm 2 nm EML激光器具备应用于50G-PON时分三模系统的技术可行性。

2.2 上行时分复用多速率接收机接收性能

在50G-PON三代时分共存系统中,上行时分复用1.25 Gbit/s、10.312 5 Gbit/s、24.883 2 Gbit/s、49.766 4 Gbit/s四速率接收机是实现多代PON共存与速率兼容的关键组件。该接收机能够支持EPON(收窄/非收窄)、10G-EPON(对称/非对称)和50G-PON(对称/非对称)三代PON 6类终端的4种上行速率,并在同一接收机内完成光电转换与信号放大,是OLT实现统一DBA调度并正确解析不同制式上行PON信号的物理层基础。上行时分复用多速率接收机主要包括雪崩光电二极管(APD)与四速率突发跨阻放大器(TIA)芯片。相较于传统EPON和10G-EPON中使用的1.25 Gbit/s与10.312 5 Gbit/s双速率时分复用突发接收机,50G-PON四速率接收机所需支持的速率等级及接收光功率范围跨度更大,因此需对其中APD芯片的带宽、响应度与倍增因子,以及突发TIA的带宽、跨阻增益、等效输入噪声与饱和电流等关键参数进行明确定义与研究^[13-14]。

基于现有10G-EPON商用OLT光模块的器件选型,为同时满足1.25 Gbit/s与10.312 5 Gbit/s速率的接收灵敏度要求,APD倍增后的响应度应大于8 A/W,且APD与TIA的总带宽需分别控制在1 GHz与7.5 GHz。对于25 Gbit/s与50 Gbit/s速

率,则要求APD与TIA的总带宽不低于18.75 GHz。由于当前APD芯片缺乏快速的带宽与增益切换机制,其各项指标必须兼顾所有速率等级的要求。因此,多速率时分复用接收机在低速模式下的带宽控制需通过突发TIA的档位切换来实现。除带宽控制外,突发TIA还需针对不同速率的输入信号,动态调整其跨阻增益。基于上述的一些关键参数,采用图3(b)所示的上行多速率突发接收性能仿真链路进行仿真。低速信号通过时钟数据恢复(CDR)模块完成数据恢复与误码判决;高速信号则借助模数转换(ADC)模块与DSP进行采集、处理与数据恢复,最终得到50G-PON四速率时分复用接收机在各速率等级下的接收灵敏度仿真结果^[15]。

图5中,1.25 Gbit/s、10.312 5 Gbit/s、24.883 2 Gbit/s和49.766 4 Gbit/s速率的接收灵敏度为-33.5 dBm@BER 10^{-12} 、-31.75 dBm@BER 10^{-3} 、-29.44 dBm@BER 10^{-2} ,以及-26.35 dBm@BER 10^{-1} 。各速率等级接收灵敏度均符合前文所述光接口指标要求,并具备一定的余量。目前产业界锗硅材料体系和铟铝砷材料体系的25G APD或50G APD可满足仿真指标要求,配合多速率模式突发TIA,上行时分复用四速率接收机具备技术可行性。

3 50G-PON三代时分复用Combo光模块及系统验证

3.1 50G-PON时分复用三模Combo光模块

50G-PON时分复用三模Combo光模块的结构如图6所示。该模块由下行方向的3个发射机——EPON 1490 nm直接调制激光器(DML)、10G-EPON 1577 nm EML和50G-PON 1366 nm EML-SOA,以及一个上行时分复用四速率突

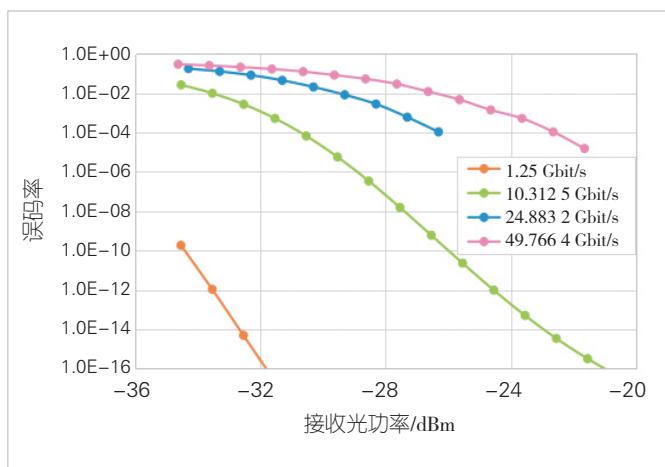


图5 上行四速率接收机接收灵敏度仿真结果

发接收机组成。模块还包含EPON与10G-EPON专用的LDD驱动芯片和突发限幅放大器(LA)芯片、50G-PON专用的LDD驱动芯片、突发线性放大器(Linear Amplifier)芯片和高速DSP芯片等电学部分。

上行时分复用四速率突发接收机内置高带宽、高灵敏度的APD芯片以及四速率突发TIA。下行3个波长通道与整个上行1 260~1 360 nm波长通道通过波分复用器件进行合波与分波，最终经由光模块的光纤适配接口实现单纤的双向传输。

与50G-PON波分复用三模Combo光模块相比^[1]，本方案无须采用复杂光路对上行1 260~1 280 nm、1 284~1 288 nm和1 290~1 330 nm 3个波长进行波分复用，有利于降低50G-PON上行链路的接收损耗，节省模块内部布局空间，从而实

现小型化的三模Combo光收发组件，并可支持16端口槽位密度的双倍密度小型可插拔(SFP-DD)封装形态。由于仅需一个突发接收机，该方案显著减少了上行突发接收机及相关配套电芯片的数量，有助于降低光模块的整体成本^[16]。

在核心光电芯片方面，50G-PON时分复用三模Combo光模块可与波分复用方案共享绝大部分芯片，例如1 577 nm激光器、1 490 nm激光器、25G/50G APD、50G-PON专用驱动放大芯片及DSP芯片等，仅需将1 342 nm EML-SOA激光器替换为1 366 nm EML-SOA。这一设计最大限度地复用了现有PON产业链资源。

3.2 50G-PON时分复用三模Combo光模块系统验证

基于前述上、下行关键光电芯片、50G-PON时分复用三模Combo光模块结构及系统架构，本文完成了包含50G-PON时分三模Combo光模块与系统设备的整套50G-PON三代时分共存系统验证，具体如图7所示。验证内容涵盖光模块的光接口指标以及系统的上下行功率预算与吞吐量性能。

实验验证结果表明，该50G-PON时分三模Combo光模块的下行发射指标均满足前文所述光接口指标要求。其中，1 366 nm通道的平均发射光功率大于9.3 dBm，中心波长为1 366.55 nm，边模抑制比优于50 dB。在BtB传输条件下，其眼图张开度良好，消光比为9.387 dB，眼图余量超过20%。经20 km光纤传输后，受色散影响，1 366 nm发射机的眼图出现一定闭合，测得的TDEC为3.17 dB。根据平均发射光功率与消光比，折算得到光模块光口的OMA-TDEC指标为8.2 dBm。该光模块其他下行通道的指标为：10G-EPON 1 577 nm通道的发射光功率为4.5 dBm，消光比为

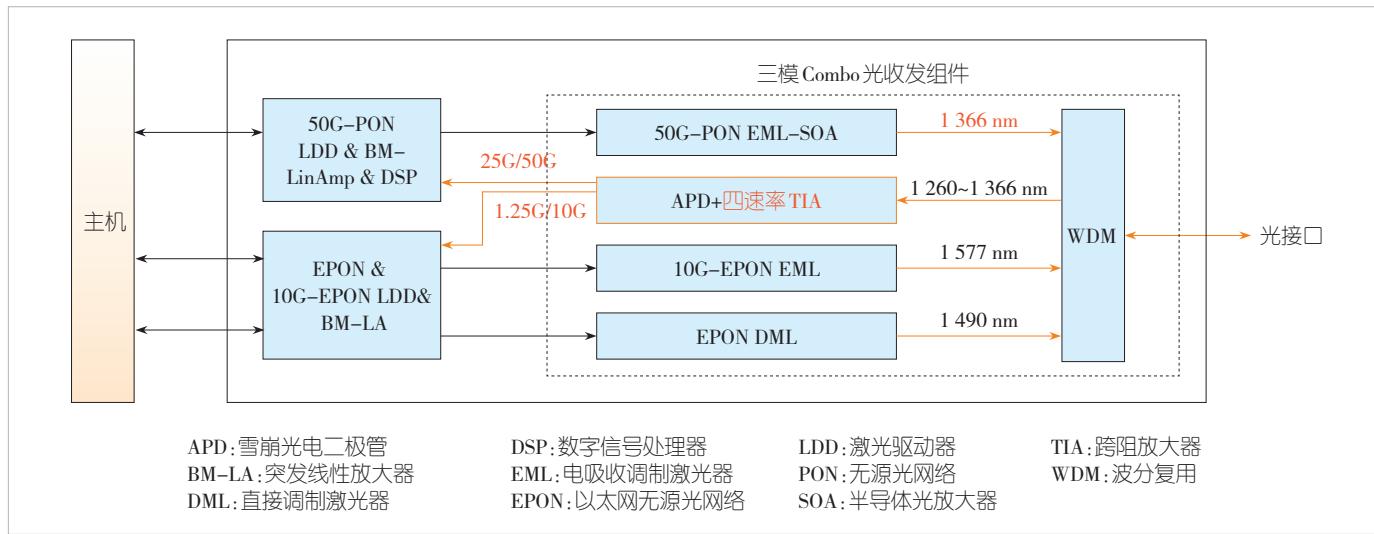


图6 50G-PON时分复用三模Combo光模块结构框图

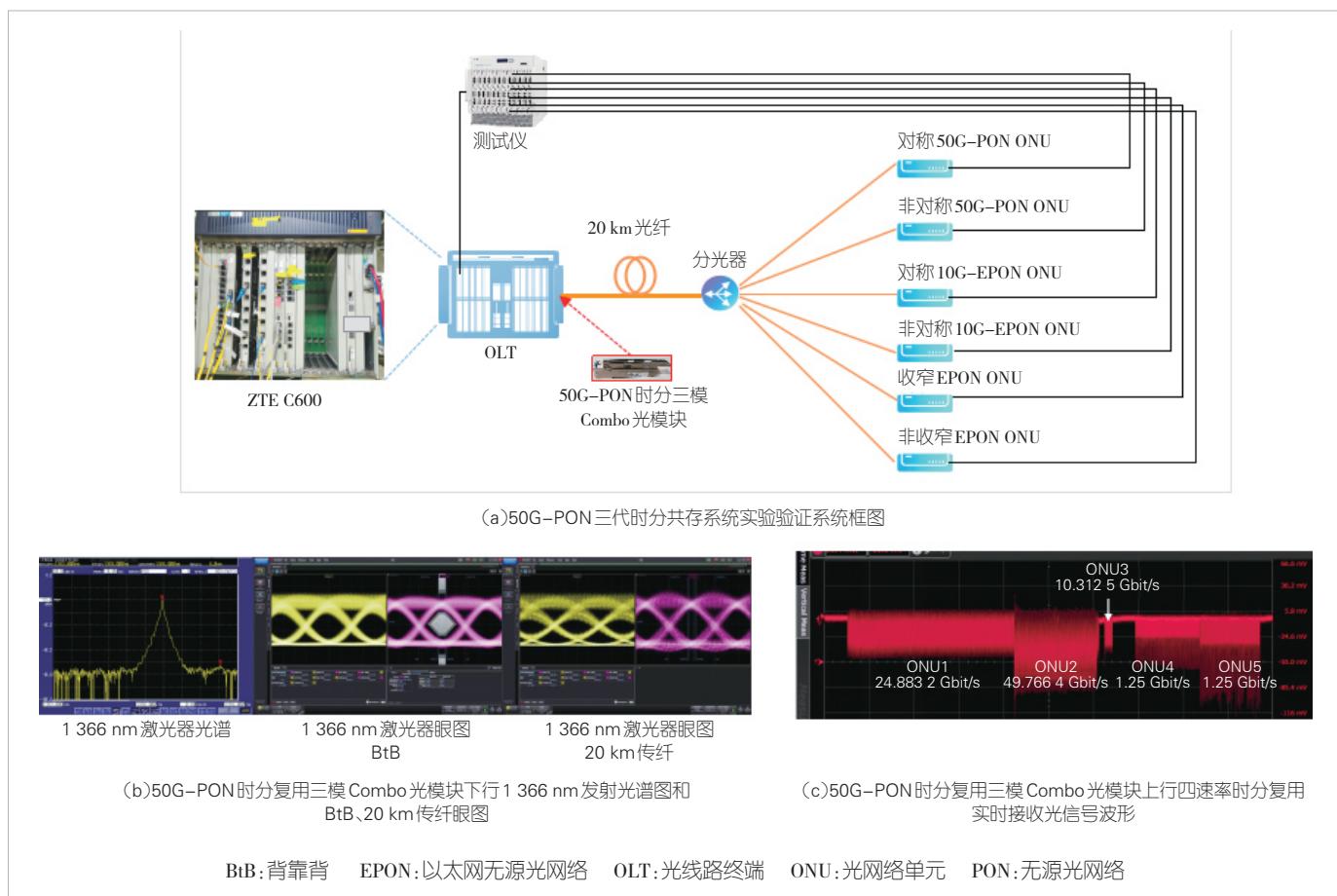


图7 50G-PON三代时分共存系统实验验证系统

7.17 dB; EPON 1490 nm 通道的发射光功率为 6.0 dBm, 消光比为 8.88 dB。

在上行方向, 50G-PON 时分三模 Combo 光模块各速率的离线误码接收灵敏度的具体实验验证结果如图 8 所示。1.25 Gbit/s 接收灵敏度为 -34 dBm @ BER 10^{-12} , 10.3125 Gbit/s 接收灵敏度为 -32 dBm @ BER 10^{-3} , 24.883 2 Gbit/s 接收灵敏度为 -28 dBm @ BER 10^{-2} , 49.766 4 Gbit/s 接收灵敏度为 -25.2 dBm @ BER 10^{-2} , 符合所述光接口指标和仿真预期。

基于图 7 所示的实时验证系统, 本文进一步对 50G-PON 时分三代共存系统的光功率预算、上下行吞吐量及长期稳定性进行了测试验证。在该实时系统中, 收窄/非收窄 EPON、对称/非对称 10G-EPON 以及对称/非对称 50G-PON 共 6 类 ONU 终端接入同一 ODN, 并通过上行混合 DBA 调度机制实现所有终端的带宽分配与稳定运行。系统中各通道的接收灵敏度定义为业务流零丢包时对应的接收光功率。测试结果表明, EPON、10G-EPON 与 50G-PON 通道的光功率预算均不低于 29 dB, 不仅满足现网部署的功率预算要求, 还保留了一定余量, 具体结果见表 1。

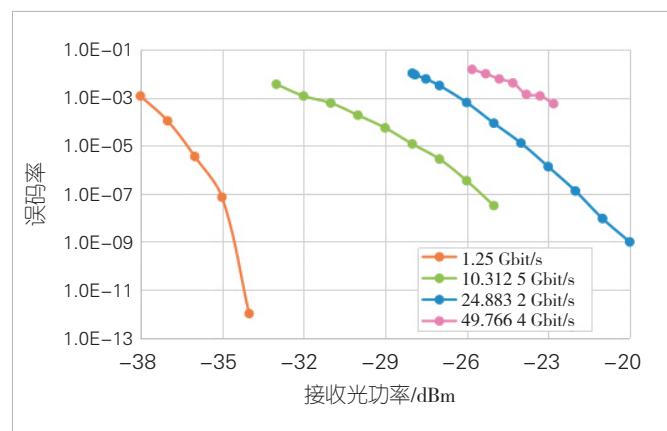


图8 50G-PON时分复用三模Combo光模块上行四速率收灵敏度实验验证结果

在 ODN 6 类 ONU 终端业务并发测试中, 50G-PON 通道在下行方向和上行方向最大业务吞吐量分别达到 40.5 Gbit/s 和 40 Gbit/s, 与 50G-PON GPON 体系波分三模方案吞吐量性能基本相当。在长期稳定性测试时, 6 类终端均 24 h 在线, 且极限带宽跑流上、下行 24 h 零丢包。

表1 50G-PON三代时分共存系统功率预算验证结果

	发射光功率/ dBm	传纤后接收 灵敏度/dBm	链路 预算/dB
50G-PON下行	9.30	-24.50	33.80
50G-PON上行非对称	7.81	-27.58	35.39
50G-PON上行对称	9.96	-24.36	34.32
10G-EPON下行	4.50	-31.94	36.44
10G-EPON上行非对称	1.81	-31.41	33.22
10G-EPON上行对称	6.32	-30.39	36.71
EPON下行	6.00	-29.30	35.30
EPON非收窄上行	1.65	-30.20	31.85
EPON收窄上行	1.30	-31.35	32.65

EPON:以太网无源光网络 PON:无源光网络

4 结束语

针对EPON与10G-EPON向50G-PON演进过程中,因兼容现网非收窄EPON终端而面临的波长冲突、终端退网难度大、成本高等问题,本文提出了50G-PON三代时分共存系统架构。该方案在下行方向采用 $1.366 \pm 2 \text{ nm}$ 新波长,并结合统一的DBA调度机制实现带宽按需分配,从而彻底解决了上下行波长冲突问题。对下行新波长的传输代价以及上行四速率(1.25 Gbit/s、10.312 5 Gbit/s、24.883 2 Gbit/s、49.766 4 Gbit/s)时分接收灵敏度等关键技术开展了深入研究。基于上述研究,完成了50G-PON三代时分共存系统设备及相关Combo光模块的研制与测试验证。测试结果表明,系统各项指标符合预期,能够满足现网ODN 29 dB光功率预算及业务吞吐量要求。本研究有效解决了运营商在EPON体系网络升级中面临的终端退网困难、多代共存波长冲突等实际问题,为GPON与EPON体系向50G-PON的平滑融合演进,提供了一条高效、低成本且可持续的技术路径。

参考文献

- [1] 黄新刚,杨波.50G-PON标准进展及关键技术[J].中兴通讯技术,2024,30(3):72-80
- [2] 张伟良,黄新刚,马壮.基于专用激活波长的低时延50G-PON原理与实现[J].中兴通讯技术,2022,28(4):58-62
- [3] ITU. Higher speed passive optical networks - requirements: G.9804.1 [S]. 2019
- [4] ITU. Higher speed passive optical networks - common transmission convergence layer specification: G.9804.2 [S]. 2021
- [5] ITU. 50-Gigabit-capable passive optical networks (50G-PON): physical media dependent (PMD) layer specification: G.9804.3 [S]. 2021
- [6] ITU. Transmission systems and media, digital systems and networks [EB/OL]. [2025-11-12]. <https://www.itu.int/rec/T-REC-G/6>
- [7] IEEE. IEEE standard for ethernet amendment 9: physical layer specifications and management parameters for 25 Gb/s and 50 Gb/s passive optical networks [EB/OL]. [2025-11-14]. <https://ieeexplore.ieee.org/document/STDGTSU24221>
- [8] 中国电信集团有限公司.中国电信光接入网技术要求EPON (1G-EPON/10G-EPON)设备: Q/CT 2361.1-2021 [S]. 2021
- [9] ITU. Coexistence of passive optical network system: G.9805 [S]. 2022
- [10] ITU. 40-Gigabit-capable passive optical networks (NG-PON2): general requirements: G.989.1 [S]. 2013
- [11] ITU. Spectral grids for WDM applications: CWDM wavelength grid: G.694.2 [S]. 2003
- [12] CHENG Y, PAN J, WANG Y, et al. 40-Gb/s low chirp electroabsorption modulator integrated with DFB laser [EB/OL]. [2025-11-06]. <https://ieeexplore.ieee.org/document/4745770>
- [13] 王新余,孔雪,贺峰.新型家庭全光网技术[J].中兴通讯技术,2023,27(2):56-62. DOI: 10.12142/ZTETJ.202301011
- [14] 刘刚,张德智,陈爱民.光接入网络遥测技术的分析与实践[J].中兴通讯技术,2023,27(2):73-78. DOI: 10.12142/ZTETJ.202303013
- [15] 陆源,牛文林,王永奔,等.基于数字子载波和概率整形的相干光通信系统设计及应用[J].中兴通讯技术,2023,27(2):78-82. DOI: 10.12142/ZTETJ.202304014
- [16] 沈百林,王会涛.相干光收发器件未来技术演进[J].中兴通讯技术,2023,27(2):82-86. DOI: 10.12142/ZTETJ.202305013

作者简介



杨波,中兴通讯股份有限公司技术预研资深专家,高级工程师;主要从事下一代PON技术预研和标准化工作;参与多项国家和省部级重点项目,主持上海市科技启明星项目1项,曾获中国电子学会科学技术进步奖一等奖;获授权专利40余项。



张德智,中国电信集团光接入领域高级专家、中国电信研究院网络技术研究所接入网络研究中心总监、国际标准组织FSAN现任主席;主要研究方向为下一代高速PON、工业PON、全光组网FTTR、光接入网SDN智能化等;长期在SG15、CCSA TC6从事标准化工作,牵头20余项ITU-T国际标准及多项光接入网领域行业标准的制定;发表论文10余篇,拥有中国专利50余项。



刘波,中兴通讯股份有限公司光接入产品线预研总工;主要从事PON技术预研方面的工作。



数字副载波复用光纤通感融合系统干涉衰落抑制研究

Interference Fading Suppression in Fiber-Optic Integrated Sensing and Communication Systems Using Digital Subcarrier Multiplexing

张兵兵/ZHANG Bingbing¹, 向梦/XIANG Meng^{1,2,3},
张泽彬/ZHANG Zebin¹, 付松年/FU Songnian^{1,2,3},
秦玉文/QIN Yuwen^{1,2,3}

(1. 广东工业大学, 中国广州 510006;
2. 广东工业大学通感融合光子技术教育部重点实验室, 中国广州 510006;
3. 广东省信息光子技术重点实验室, 中国广州 510006)
(1. Guangdong University of Technology, Guangzhou 510006, China;
2. Key Laboratory of Photonic Technology for Integrated Sensing and Communication, Ministry of Education of China, Guangzhou 510006, China;
3. Guangdong Provincial Key Laboratory of Information Photonics Technology, Guangzhou 510006, China)

DOI: 10.12142/ZTETJ.202506005

网络出版地址: <https://link.cnki.net/urlid/34.1228.tn.20251218.1454.006>

网络出版日期: 2025-12-19

收稿日期: 2025-11-03

摘要: 针对数字副载波复用光纤通感融合 (DSM-ISAC) 系统中相干探测型相位敏感光时域反射仪 (Φ -OTDR) 因干涉衰落导致传感性能劣化和可靠性下降的挑战, 提出了一种接收端多域集成数字信号处理衰落抑制方案。通过在接收端引入相位域的相移变换 (PST)、频率域的脉冲内频分 (IFD) 和空间域的移动旋转矢量平均 (MRVA), 实现对干涉衰落的协同抑制。为了验证所提方案的有效性, 搭建了 DSM-ISAC 实验平台, 利用 DSM 信号的频谱灵活分配特点来优化子载波间保护间隔, 在同一波长信道中生成 36 GBaud DP-16QAM DSM 通信信号与线性调频 (LFM) 脉冲传感信号, 并经过 920 m 标准单模光纤传输, 完成了对系统干涉衰落抑制的实验验证。与传统干涉衰落抑制方案相比, 所提出的 PST-IFD-MRVA 方案可显著抑制干涉衰落, 提高强度信噪比近 10 dB 且增强应变灵敏度到 $9.09 \text{ p}\varepsilon/\text{Hz}$, 成功解调出施加于光纤 820 m 处的 10 kHz 扰动信号。实验结果表明该研究方案在提升 DSM-ISAC 系统的传感可靠性方面具有可行性。

关键词: 光纤通感融合系统; 数字副载波复用; 相位敏感光时域反射仪; 干涉衰落抑制

Abstract: As for fiber-optics integrated sensing and communication systems enabled by digital subcarrier multiplexing (DSM-ISAC), the coherent-detection based phase-sensitive optical time-domain reflectometer (Φ -OTDR) suffers from impaired sensing performance and insufficient reliability due to the interference fading effect. A receiver-side multi-domain integrated fading suppression scheme is proposed. Our proposed scheme sequentially applies phase shift transformation (PST) in the phase domain, inner-pulse frequency division (IFD) in the frequency domain, and moving rotation vector averaging (MRVA) in the spatial domain to achieve joint suppression of interference fading. To verify its performance, a DSM-ISAC experimental platform is established. Leveraging the flexibility of spectral allocation in DSM signals, a 36 GBaud DP-16QAM DSM communication signal and a linear frequency modulated (LFM) pulse sensing signal are generated within the same wavelength channel and co-propagated over 920 m of standard single-mode fiber (SSMF), with an optimized protection interval bandwidth between the two central subcarriers. Experimental results indicate that, in comparison with conventional suppression schemes, the proposed PST-IFD-MRVA scheme significantly enhances fading suppression performance, achieving an improvement in signal-to-noise ratio of nearly 10 dB and a strain sensitivity of $9.09 \text{ p}\varepsilon/\text{Hz}$. Meanwhile, a 10 kHz vibration signal applied at the position of 820 m of SSMF is successfully demodulated. These findings validate the feasibility and promising potential of the proposed scheme in enhancing the sensing reliability of DSM-ISAC systems.

Keywords: fiber-optics integrated sensing and communication; digital subcarrier multiplexing; phase-sensitive optical time-domain reflectometer; interference fading suppression

引用格式: 张兵兵, 向梦, 张泽斌, 等. 数字副载波复用光纤通感融合系统干涉衰落抑制研究 [J]. 中兴通讯技术, 2025, 31(6): 29–38. DOI: 10.12142/ZTETJ.202506005

Citation: ZHANG B B, XIANG M, ZHANG Z B, et al. Interference fading suppression in fiber-optic integrated sensing and communication systems using digital subcarrier multiplexing [J]. ZTE technology journal, 2025, 31(6): 29–38. DOI: 10.12142/ZTETJ.202506005

光纤通信系统作为现代信息社会的基础承载设施，凭借低损耗、超宽带和抗干扰等优越传输特性，在全球范围内得到广泛部署。目前光纤光缆铺设总长度已超过 5×10^8 km，构建了覆盖全球的通信网络。与此同时，分布式光纤传感技术不仅赋予光纤信息传输能力，还能对温度、振动、应变等环境参数进行高精度分布式监测，具备对外界扰动的实时感知与智能监控能力^[1]。由于高速光纤通信与分布式光纤传感在信号调制、复用和探测方式等方面具有高度相似性，利用同一根光纤实现兼具通信与感知功能的通感融合（ISAC）系统逐渐成为全球研究热点，被视为信息光子技术的重要发展方向。

现有光纤通感融合系统主要包括两类：基于前向通信信号的通感融合系统和基于瑞利背向散射（RBS）^[4]的通感融合系统。前者利用前向传输信号完成通信功能，并通过相干接收解调相位^[2]和偏振态^[3]信息实现环境扰动感知，但该方案依赖于整条光纤链路的相位累积或偏振演化信息，难以实现多扰动事件的同时定位，且定位精度有限。相比而言，基于RBS^[4]的通感融合系统通常采用相位敏感光时域反射仪（Φ-OTDR）实现环境扰动定位，能够区分同时发生的多个事件，具备定位精度高、感知灵敏度高、单端接入和无需时间同步等性能优势。为降低通信与感知信号间的干扰并提升频谱利用率，基于RBS的共波长信道数字副载波复用光纤通感融合（DSM-ISAC）^[5-6]架构近年来受到广泛关注。

相干探测^[7]是长距离感知的关键技术。本振光（LO）与RBS混频后，可提高RBS信噪比，增强感知灵敏度。然而，相干探测同时面临偏振衰落和干涉衰落的问题。偏振衰落源于LO与RBS的偏振态失配，目前业界通常采用偏振分集接收技术^[8]进行抑制。干涉衰落则源于光纤的非均匀折射率分布对RBS相位的随机调制^[9]，当大量散射体产生的RBS发生干涉相消时，会形成信号衰落点。在这些低信噪比区域，解调出相位信息误差大，扰动信息失真，继而引发ISAC系统漏报或误报，严重限制了系统可靠性。

基金项目：国家自然科学基金项目（62435004、U21A20506、62475051）；广东省“珠江人才计划”引进创新创业团队项目（2021ZT09X044）

本文在数字副载波复用光纤通感融合系统框架下，针对相干探测Φ-OTDR的干涉衰落问题开展了干涉衰落抑制研究，提出了一种接收端多域集成数字信号处理衰落抑制方案，并进行了系统参数优化，最终完成了实验验证和衰落抑制性能对比分析。

1 数字副载波复用光纤通感融合系统

DSM技术不仅因其子载波的灵活可调而天然适用于信号复用，还对光纤非线性效应及均衡增强相位噪声（EPPN）^[10]表现出高容忍度，从而保障了通信链路的稳定性和鲁棒性。在传感方面，为克服传统Φ-OTDR中空间分辨率与脉冲宽度之间的性能制约，采用线性调频脉冲（LFM）结合匹配滤波（MF）脉冲压缩技术，以同时实现高空间分辨率与长距离探测能力^[11]。

1.1 通感融合信号产生

如图1所示，DSM通信信号由多个灵活可调的子载波构成，在中间相邻两个子载波之间预留保护间隔，并插入窄带传感信号，从而实现共波长信道的通感融合信号一体化。发射端DSM通信信号的电场表达式为：

$$S_{\text{DSM}}(t) = \sum_{k=1}^{N_{\text{sc}}} s_k(t) \cdot \exp(j(2\pi f_k t)) \quad (1),$$

其中， N_{sc} 为子载波数目， $s_k(t)$ 为第 k 个子载波基带信号， f_k 为第 k 个子载波中心频率。

发射端产生的调频脉冲传感信号 $S_{\text{CP}}(t)$ 可表示为：

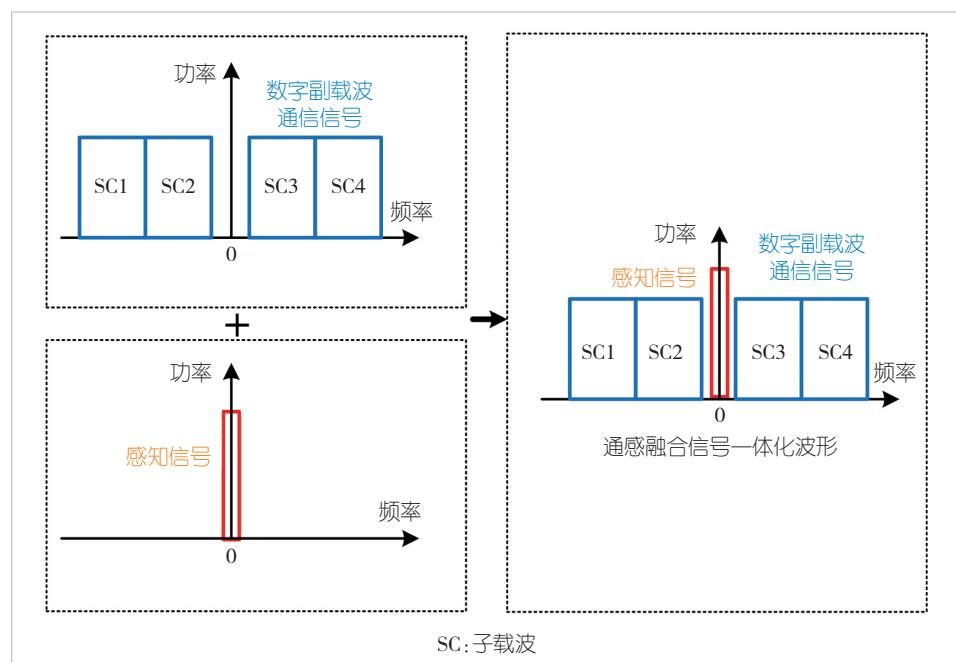


图1 基于数字副载波复用的通感融合信号产生

$$S_{\text{CP}}(t) = \exp j(2\pi f_0 t + \pi k t^2) \cdot \text{rect}\left(\frac{t}{\tau_p}\right) \quad (2),$$

其中, f_0 为起始频率, $k = B/\tau_p$ 为调频速率, B 为调频带宽, τ_p 为脉冲持续时间。

经过频分复用后可以得到通感融合光信号 $S_{\text{ISAC}}(t)$ 的表达式为:

$$S_{\text{ISAC}}(t) = [S_{\text{CP}}(t) + S_{\text{DSM}}(t)] \cdot \exp j(2\pi f_c t) \quad (3),$$

其中, f_c 为光载波中心频率。

1.2 通感融合信号接收

经过单模光纤前向传输后, 通感融合信号首先进入高速相干接收机, 通过数字域窄带滤波完成通感信号分离, 最后通过数字均衡, 恢复出通信信号。携带了外界扰动信息的 RBS 光信号经环形器到达传感信号接收端。背向 ISAC 信号经过滤波器滤除通信信号以及放大自发辐射 (ASE) 噪声, 得到调频脉冲对应的背向散射信号。在相干探测方案中, RBS 表示为:

$$E_{\text{RBS}}(t) = \sum_{i=1}^N r(i) \exp\left(-\alpha \frac{c\tau_i}{n}\right) \exp\left[j[2\pi(f_c + f_0)(t - \tau_i) + \pi k(t - \tau_i)^2]\right] \text{rect}\left(\frac{t - \tau_i}{T_p}\right) \quad (4),$$

其中, N 为整个光纤链路的散射点数, $r(i)$ 为瑞利散射系数, α 为光纤损耗系数, τ_i 为光从第 i 个散射点返回到传输接收端的飞行时间, c 为真空中的光速, n 为纤芯折射率, 本振光的光场表达式为:

$$E_{\text{LO}}(t) = \sqrt{P_{\text{LO}}} \exp j(2\pi f_c t) \quad (5).$$

背向散射光与本振光拍频后的电信号表达式为:

$$I(t) = R \sum_{i=1}^N r(i) \exp\left(-\alpha \frac{c\tau_i}{n}\right) \cos[2\pi f_0(t - \tau_i) + \pi k(t - \tau_i)^2 - 2\pi f_c \tau_i] \quad (6),$$

其中, R 为光电探测器的响应度, 在数字信号处理中, 对 $I(t)$ 作希尔伯特变换可得到复数拍频信号, 其表达式为:

$$\begin{aligned} I(t) &= R \sum_{i=1}^N r(i) \exp\left(-\alpha \frac{c\tau_i}{n}\right) \exp j\left[\frac{2\pi f_0(t - \tau_i) + \pi k(t - \tau_i)^2 - 2\pi f_c \tau_i}{\pi k(t - \tau_i)^2 - 2\pi f_c \tau_i}\right] \\ \text{rect}\left(\frac{t - \tau_i}{T_p}\right) &= \int_0^{T_p} R r \cdot \exp\left(-\alpha \frac{c\tau}{n}\right) \exp j\left[\frac{2\pi f_0(t - \tau) + \pi k(t - \tau)^2 - 2\pi f_c \tau}{\pi k(t - \tau)^2 - 2\pi f_c \tau}\right] d\tau \quad (7), \\ \text{rect}\left(\frac{t - \tau}{T_p}\right) d\tau &= R [h(t) \otimes S_{\text{CP}}(t)] \end{aligned}$$

其中, T_p 表示最大往返时间, $S_{\text{CP}}(t)$ 为探测光信号, 公式

(7) 可以表示为探测光信号 S_{CP} 与光纤冲击响应 $h(t)$ 的卷积形式, 即 $I(t) = R \cdot [h(t) \otimes S_{\text{CP}}(t)]$, 其中 \otimes 表示卷积。这时的空间分辨率由脉冲持续时间决定。为了提高空间分辨率, 对拍频信号进行匹配滤波处理:

$$\begin{aligned} I_{\text{MF}}(t) &= R \cdot [h(t) \otimes S_{\text{CP}}(t) \otimes S_{\text{CP}}^*(-t)] = \\ &R \cdot [h(t) \otimes P(t)] \end{aligned} \quad (8),$$

其中, $S_{\text{CP}}^*(-t)$ 为匹配滤波器, $P(t)$ 为压缩后的脉冲信号, 然后求解差分相位, 可以恢复出外界扰动信号。

2 干涉衰落抑制

干涉衰落作为 Φ -OTDR 的重要研究挑战, 已有多种衰落抑制方案被提出并验证。其中, 频率分集 (FD) 方案利用不同频率脉冲在空间上具有独立衰落特性, 已成为主流技术路线^[12]。此外, 波长分集 (WD)^[13] 与相位分集 (PD)^[14] 等方案亦被证明能够有效产生相互独立的信号分量。这类分集方案的共同特征在于产生并合成多个独立的 RBS, 以抑制在特定光纤位置处的信号衰落。然而, 由于这些分量均为复数信号, 且相位各不相同, 直接叠加可能因相位抵消而无法增强信号。为此, 研究人员提出了旋转矢量和 (RVS) 方案^[15], 通过先对各分路信号进行相位旋转以实现相位对齐, 再进行矢量叠加, 从而有效消除信号衰落点, 并利用矢量平均降低噪声, 提高信噪比。尽管上述分集技术具有良好的抑制效果, 但其实现往往依赖复杂器件或特殊光纤, 导致系统复杂度和成本显著增加。鉴于此, 如何在不增加系统硬件复杂度的前提下实现高效的干涉衰落抑制, 成为前沿研究方向。其中, 基于接收端数字信号处理的干涉衰落抑制方案因具备灵活性与成本优势而受到广泛关注。

2.1 脉冲内频分法

为将频率分集思想引入基于 LFM 的 Φ -OTDR 系统, 文献[15]提出了脉冲内频分法 (IFD) 以抑制干涉衰落。当 LFM 光脉冲在光纤中传播时, 其 RBS 因弹性散射特性, 能够完整保留原始调频特征。在调频带宽 F 足够大的条件下, 可利用数字滤波器将每次探测得到的 RBS 划分为多个子带信号。若将调频范围 F 划分为 M 个子带, 则每个子带的带宽为:

$$\Delta F = \frac{F}{M}, m = 1, 2, \dots, M \quad (9).$$

通过构建匹配滤波器组, 其中每个滤波器仅与原始调频脉冲的特定频谱子带相对应, 即可从单次探测的 RBS 中解调出多个经脉冲压缩的瑞利信号:

$$s_m(z) = \int r(t, z) h_m^*(t) dt \quad (10),$$

其中, $r(t, z)$ 表示在位置 z 处接收到的 RBS, $h_m^*(t)$ 表示与第 m 个子带相匹配的滤波器。脉冲压缩后的瑞利信号在形式上与单频 Φ -OTDR 获取信号一致, 其中心频率由对应的调频范围决定, 因此各子信号的中心频率各不相同, 从而具有相互独立的衰落特性。最终, 通过 RVS 算法对这些子信号进行相位对齐与矢量叠加:

$$S(z) = \sum_{m=1}^M s_m(z) e^{-j\phi_m} \quad (11)$$

一般而言, 子信号数量 M 越多, 衰落抑制效果越显著。但由于边际效应的存在, 过度增加 M 并不能无限提升衰落抑制性能, 反而会导致空间分辨率的下降。

2.2 相移变换法

相位分集的核心思想是通过对原始信号施加 π 相移, 构造一个在幅度上与原始信号互补的信号, 从而抑制干涉衰落。具体而言, 可以通过选择幅值更高的信号进行相位解调或将原始信号与互补信号通过 RVS 算法合成实现衰落抑制。为了在数字域上实现相位分集, 文献[16]提出了相移变换法 (PST), 首先对采集到的时域背向散射信号 $r(t, z)$ 进行快速傅里叶变换 (FFT), 得到频域信号 $R(f, z)$:

$$R(f, z) = Fr(t, z) \quad (12)$$

然后, 在频域中, 将信号乘以特定符号函数 $H(f)$,

$$\tilde{R}(f, z) = R(f, z) \cdot H(f),$$

$$H(f) = \begin{cases} -1, & f \in \text{选定子频段} \\ 1, & \text{其他频段} \end{cases} \quad (13)$$

该操作相当于对选定频段信号施加 π 相移, 从而获得与原始信号幅度互补的子信号。随后, 通过逆傅里叶变换 (IFFT) 将频域信号转换回时域, 得到互补信号:

$$\tilde{r}(t, z) = \mathcal{F}^{-1} \tilde{R}(f, z) \quad (14)$$

单次相移变换可能无法覆盖所有信号衰落情况, 因此需要通过多次调整频域滤波函数, 生成多个相移子信号并进行优化, 以实现干涉衰落的完全抑制。在文献[16]的实验中, 通过生成并合成 6 个相移子信号, 成功消除了所有信号衰落点。然而, 增加子信号数量会显著提升计算复杂度。

2.3 移动旋转矢量平均法

在使用 RVS 算法来合成独立子带信号时, 由于各向同性噪声的随机分布, 旋转后的每个子带信号实际上并不处于完全相同的方向, 导致存在残留衰落。因此, 文献[17]提出

了移动旋转矢量平均法 (MRVA) 来抑制随机噪声。MRVA 作为一种空间域 (SSD) 的信号合成方式, 通过在一个时间窗口内, 先旋转对齐各个空间点的信号矢量以消除其固有的随机相位差, 然后进行矢量平均, 这个过程可以表示为:

$$v_{\text{SSD}}(z_m, t_i) = \sum_{\xi=1}^S \left[\left(\frac{1}{M} \sum_{k=m}^{m+M-1} A_{s, \xi}(z_k, t_i) e^{j\theta(z_m, t_i)} \right) \right] \quad (15)$$

其中, v_{SSD} 为空间域合成后的信号, S 为子信号个数, M 为窗口宽度, $A_{s, \xi}$ 为信号振幅, $e^{j\theta(z_m, t_i)}$ 为旋转因子。利用空间上相邻点的、幅度更强的信号来补偿衰落点的信号, 同时平均消除不相关噪声, 提升了衰落点的信噪比, 实现了干涉衰落的抑制。为了获得更好的衰落抑制效果, 需要增加滑动窗口的长度, 然而增加窗口长度会牺牲系统的空间分辨率, 因此在实际应用中需要做出权衡。

2.4 多域集成衰落抑制方案

现有的 IFD、PST 以及 MRVA 均是从单一维度对干涉衰落进行抑制, 各有局限性。为进一步增强信号衰落抑制效果, 本文提出了一种多域集成抑制策略, 将相位、频率和空间维度技术相结合^[18], 并将这一思想推广到基于线性调频脉冲的 DSM-ISAC 系统中, 以实现更鲁棒的干涉衰落协同抑制。需要指出的是, 由于匹配滤波和 PST 均依赖完整的信号频谱, 若先行实施 IFD 分解, 可能导致频谱截断并破坏 PST 构造的互补关系, 因此在所有集成路径中, 必须确保 PST 操作优先于 IFD 分解。基于这一原则, 本文设计了 6 种不同的算法路径, 如图 2 所示。方案 (I): 仅采用 PST 生成互补信号对, 随后利用 RVS 算法进行信号合成; 方案 (II): 利用 3 个中心频率各异的匹配滤波器将原始信号分解为 3 路子信号, 并通过 RVS 算法完成信号合成; 方案 (III): 首先经过 PST 生成互补信号对, 然后将 IFD 分解为 6 路子信号, 最后利用 RVS 算法进行合成; 方案 (IV)、(V) 和 (VI): 这 3 种方案在分解流程上与方案 (I)、(II)、(III) 相同, 但在信号合成阶段采用 MRVA 算法实现空间域合成。其中, 方案 (VI) 实现了相位、频率和空间域的级联抑制, 在理论上可达到最佳的衰落抑制效果, 但相应的计算复杂度也最高。在实际应用中, 可根据系统需求和应用场景灵活选择不同的干涉衰落抑制算法, 在提高衰落抑制性能的同时兼顾计算效率。这种多域集成抑制策略为 Φ -OTDR 系统在复杂干涉环境下的稳定运行提供了新的解决思路。

3 实验装置与结果

为验证多域集成衰落抑制算法在 DSM-ISAC 系统架构中

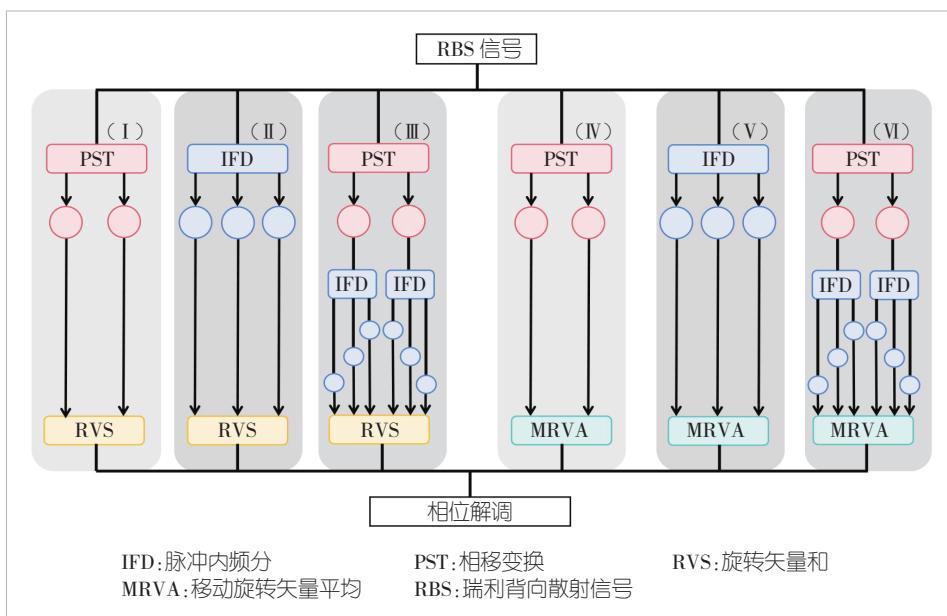


图2 多域集成信号衰落抑制实施方案

的有效性，我们搭建了实验平台，如图3所示。

3.1 实验设置

我们采用一台中心波长为1550 nm、线宽100 Hz、输出功率16 dBm的超窄线宽半导体激光器作为光源，并将输出光均分为两路，分别用于传感与通信信号的产生。在上支路中，首先利用任意函数发生器（AFG）产生线性调制电压信

号，以驱动压控振荡器（VCO）输出线性调频电信号。该电信号进一步驱动IQ调制器（IQM）产生调频光信号，经过放大与滤波后，再经声光调制器（AOM）调制为脉宽200 ns、带宽500 MHz的线性调频光脉冲，用于传感。在下支路中，首先离线生成包含4个副载波、总波特率为36 Gbaud的双偏振16阶正交幅度调制（DP-16QAM）数字副载波复用信号，并采用滚降因子为0.01的根升余弦（RRC）滤波以压缩带宽，并在中间两路副载波之间预留保护间隔（PI）。经重采样后，该信号可以加载到采样率为120 GSa/s、分辨率为8 bit的任意波

形发生器（AWG），转换为电信号后，再送入相干驱动调制器（CDM）调制生成光通信信号。最终，光通信信号与光传感信号在光纤耦合器中复用，形成ISAC信号。

ISAC信号经820 m标准单模光纤传输后，在光纤末端通过压电陶瓷换能器（PZT）施加外界扰动。PZT缠绕光纤长度为13.17 m，施加的正弦扰动频率为10 kHz。前向传输光

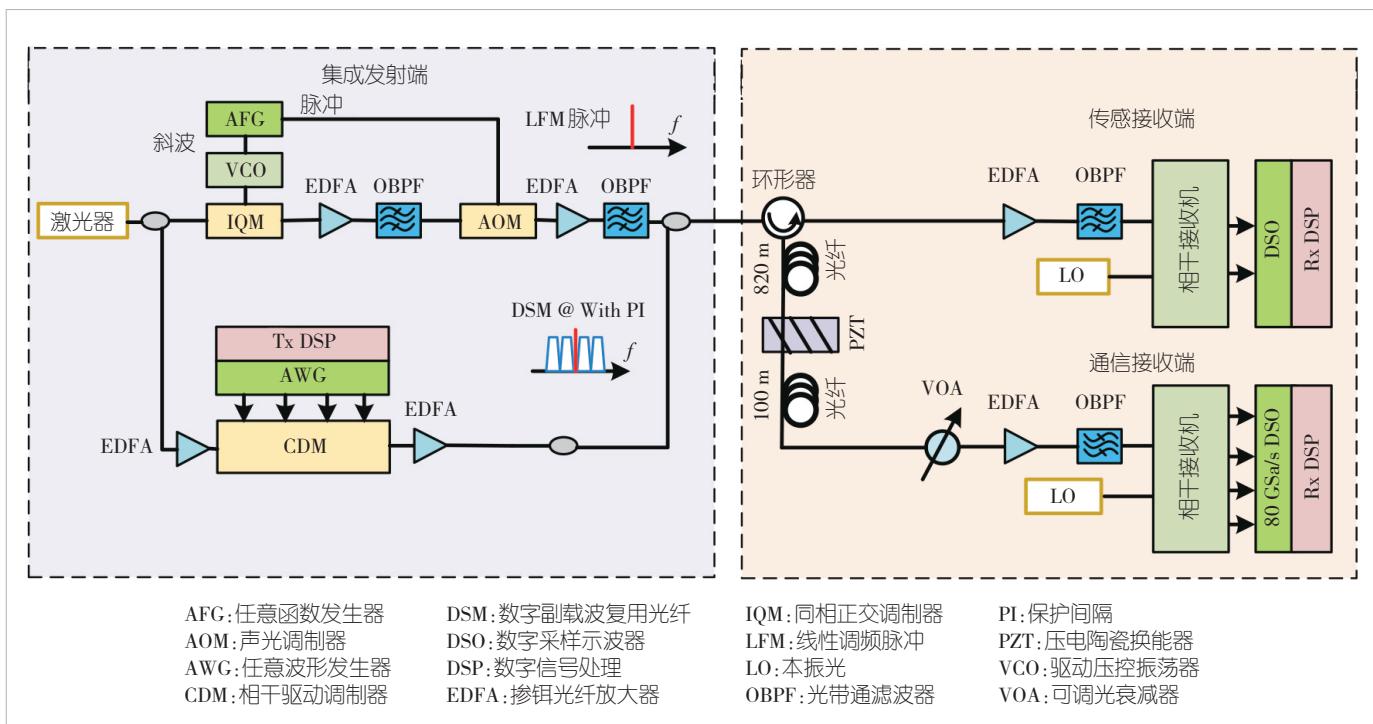


图3 光纤通感融合实验示意图

在进入接收端前经过光放大和光滤波，再与线宽为 100 kHz 的外腔激光器产生的本振光在相干接收机中进行相干检测。相干接收机输出的电信号由采样率为 80 GSa/s 的数字采样示波器 (DSO) 采集。通信接收端的数字信号处理 (DSP) 流程包括色散补偿 (CDC)、粗频偏估计与补偿 (FOC)、副载波解复用、自适应均衡、载波相位恢复以及平均误码率 (BER) 计算。在副载波解复用过程中，RRC 滤波器不仅用于副载波提取，还能有效滤除调频脉冲信号。

携带外界扰动信息的 RBS 经环形器进入传感信号接收端。背向散射光信号首先通过光滤波器去除通信分量和 ASE 噪声，得到与调频脉冲对应的背向散射分量。随后，利用相干接收机对信号进行相干检测，并由示波器采集以便在数字域进行信号处理。以 PST-IFD-MRVA 方案为例，数字信号处理流程包括 PST、IFD、MRVA 合成、相位解调、相位差分以及扰动波形重构。需要注意的是，PST 仅对信号的相位谱进行调制，而不改变幅度谱，因此生成的互补信号对需使用同一匹配滤波器以保持相位对应关系。在 IFD 处理中，两路信号分别通过同一组子带匹配滤波器。最终，选择 25 μ s 窗口长度以兼顾衰落抑制效果以及空间分辨率。

3.2 数字副载波复用光纤通感融合系统性能优化

我们首先研究了不同通信波特率下传感信号对系统传输性能的影响，实验结果如图 4 (a) 所示。实验中，DP-16QAM DSM 信号中间两副载波之间预留的保护间隔 (PI) 带

宽为 8 GHz。随着通信波特率由 30 GBaud 提升至 42 GBaud，系统误码率持续上升。更高的波特率信号占用了更宽的 DSM 信号带宽，从而加剧了由于器件带宽受限导致的码间串扰。当传感信号同时存在时，通信信号的 BER 进一步恶化，这表明引入传感功能会对通信应用带来额外损伤。

随后，我们进一步分析了不同 PI 带宽对通信性能的影响，如图 4 (b) 所示。在无传感信号时，随着 PI 由 2 GHz 增加至 10 GHz，通信信号带宽逐渐增大，受限于光电器件带宽，通信信号 Q 因子由 8.32 dB 下降至 7.20 dB。当存在传感信号时，通信信号 Q 因子由 7.70 dB 下降至 6.88 dB，相应的 Q 因子损伤则由 0.62 dB 减小至 0.32 dB。结果表明，适当增大 PI 能够有效缓解传感信号对通信信号的串扰。总体而言，传感信号的引入确实会引起通信性能的恶化。尽管通信与传感信号在频域上实现了正交复用，但光纤的非线性效应仍会使传感信号对通信信号产生非线性串扰。此外，通信接收端的数字示波器需同时采集包含传感信息的复用信号，在量化比特有限条件下，这会降低通信信号的有效分辨率，从而进一步削弱通信性能。综合考虑后，本文最终选用 36 GBaud DP-16QAM DSM 信号作为通信信号，并将 PI 设置为 8 GHz，以在通信与传感性能之间实现平衡。

3.3 衰落抑制

我们首先验证了不同方法的应变解调能力，实验结果如图 5 所示。

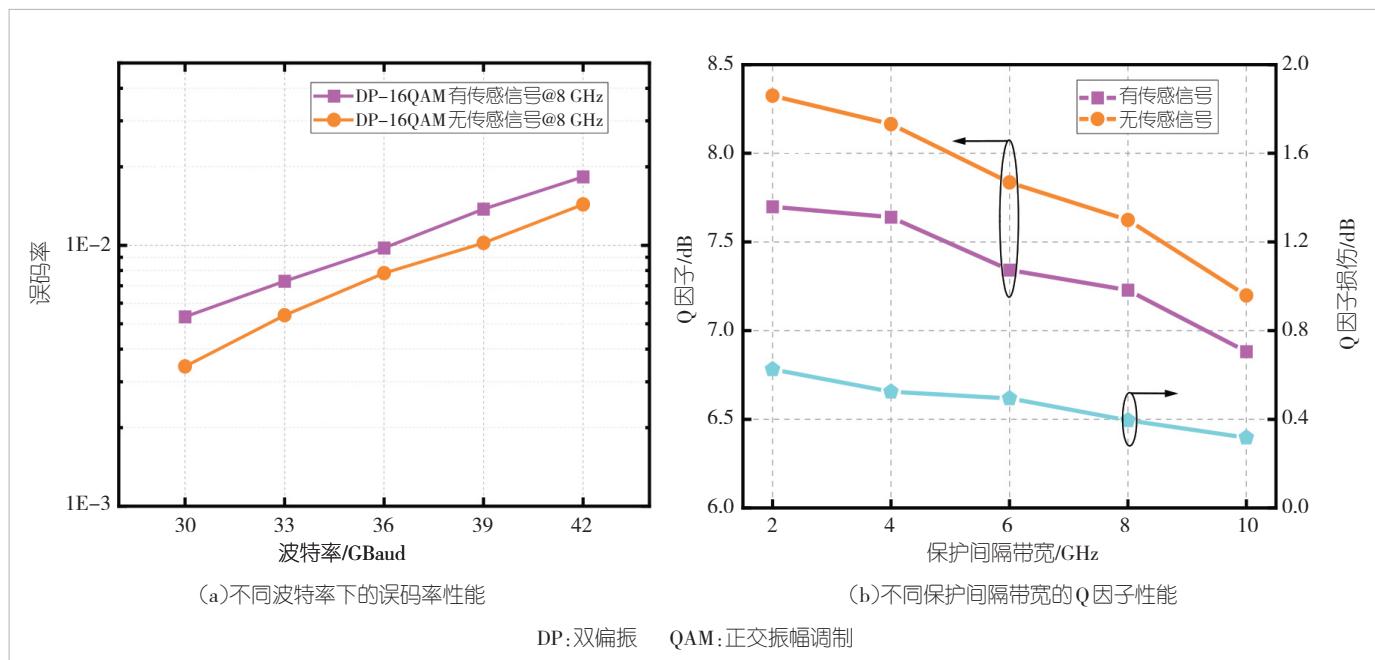


图 4 不同波特率的通信信号与保护间隔优化

图5 (a) — (f) 左图分别展示了各方案在820 m位置恢复的10 kHz正弦扰动的时域波形, 右图则给出了对应的幅度谱密度 (ASD)。可以看出, 所有方法均能够准确重构扰动信号, 且重构的时域波形无明显畸变, 说明各衰落抑制方案均具备应变跟踪能力。相应地, 在ASD中均可观察到位于10 kHz处的显著主瓣峰值, 其主瓣带宽与窗函数设定保持一致, 表明各方案对扰动的幅度和频率均能实现精确定位。

为进一步验证所提方案的衰落噪声抑制性能, 我们对比了7种处理方法, 包括6种衰落抑制方法以及不采用任何衰落抑制的原始信号处理方法。图6展示了7种不同的处理方法所得到强度、差分相位以及相位标准差 (SD) 曲线。

实验结果表明, 当未采用任何抑制手段时, 存在大量深度衰落点, 差分相位曲线出现密集波动, 对应位置的信噪比极低。这会导致相位解调严重失真, 相位标准差显著增大, 振动信号完全淹没于噪声之中。采用单一域方法时, 衰落现象得到部分缓解。其中, PST算法能够有效减少部分衰落点和相位波动, 但在非振动区域仍出现错误, 且在曲线末端难以区分振动与噪声; IFD算法则表现更优, 强度衰落点和相位波动均明显减少, 相位标准差整体下降, 仅在个别位置出现错误, 振动特征已能够被辨识。然而, 单一域方法仍难以完全消除衰落噪声。相比之下, 多域集成方法充分结合了相

位域、频域与空间域的互补优势, 使强度曲线中的多个衰落位置得到完全消除, 相位标准差曲线平滑连续, 非振动区域未出现错误, 差分相位曲线末端的振动特征亦能清晰呈现。特别是在PST-IFD-MRVA级联方案下, 衰落抑制效果最为显著, 这进一步证明了多域集成抑制方法在提升ISAC系统鲁棒性的优势。

为了定量评估不同方法的衰落抑制能力, 我们计算并比较了4项性能指标, 包括平均幅度波动、强度信噪比 (SNR)、应变灵敏度以及振动信噪比。具体而言, 平均幅度波动被定义为在非振动区域内, 单条迹线归一化幅度的最大值与最小值之比, 并对多条迹线的结果进行平均。该指标直接反映了信号衰落严重程度, 其值越小表明衰落抑制效果越好。强度SNR则用于衡量信号的稳定性, 其计算公式为:

$$\text{IntensitySNR} = 10 \log_{10} \left(\frac{\sum_z (E[A(z, t)])^2}{\sum_z \text{Var}[A(z, t)]} \right) \quad (16),$$

其中, $A(z, t)$ 表示信号幅值, $E(\cdot)$ 和 $\text{var}(\cdot)$ 分别表示时间平均与方差, $\sum_z(\cdot)$ 表示沿光纤距离的累加。该指标值越高, 说明信号强度越稳定。在评估解调性能方面, 应变灵敏度通过计算无扰动区域内应变信号ASD的平均值来确定, 用于刻画系统可探测最小应变能力。振动SNR则定义为扰动频谱中信号峰值功率与噪声基底平均功率的比值, 用来衡量系统在

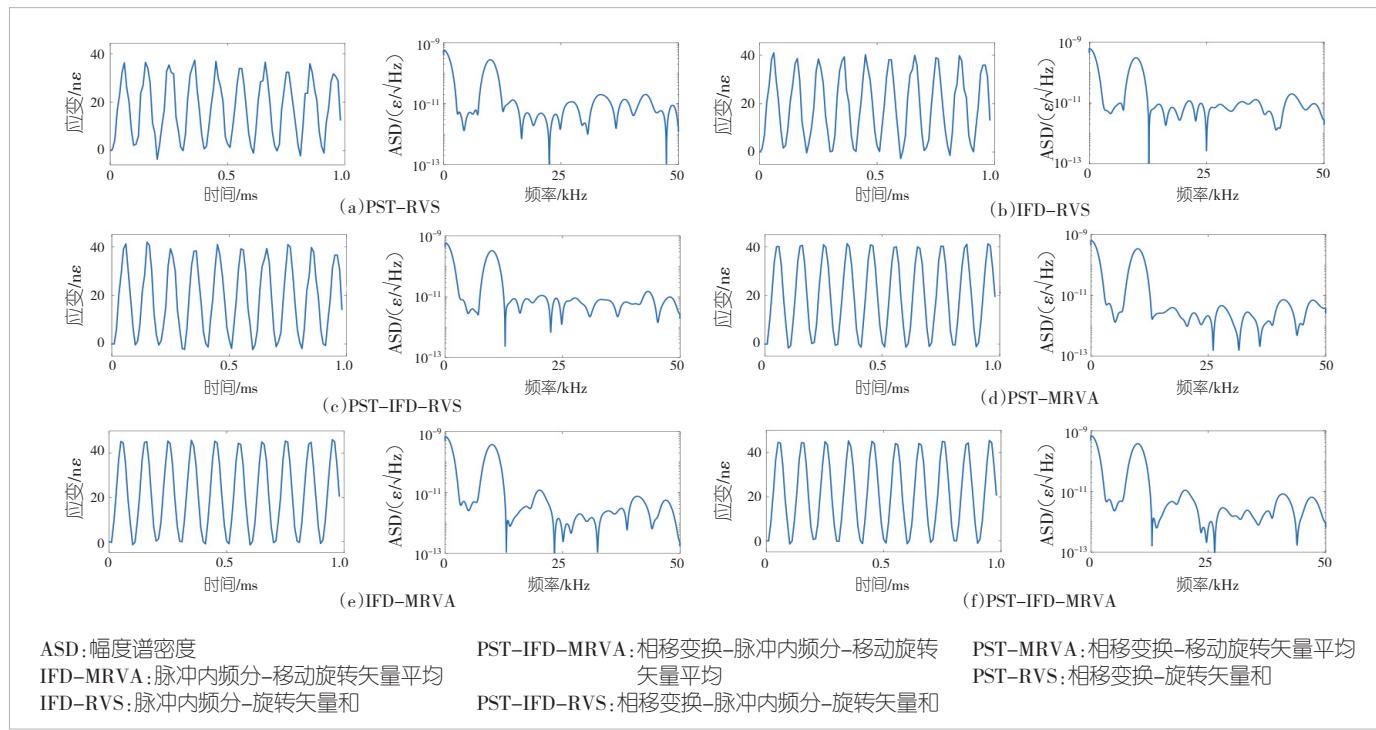
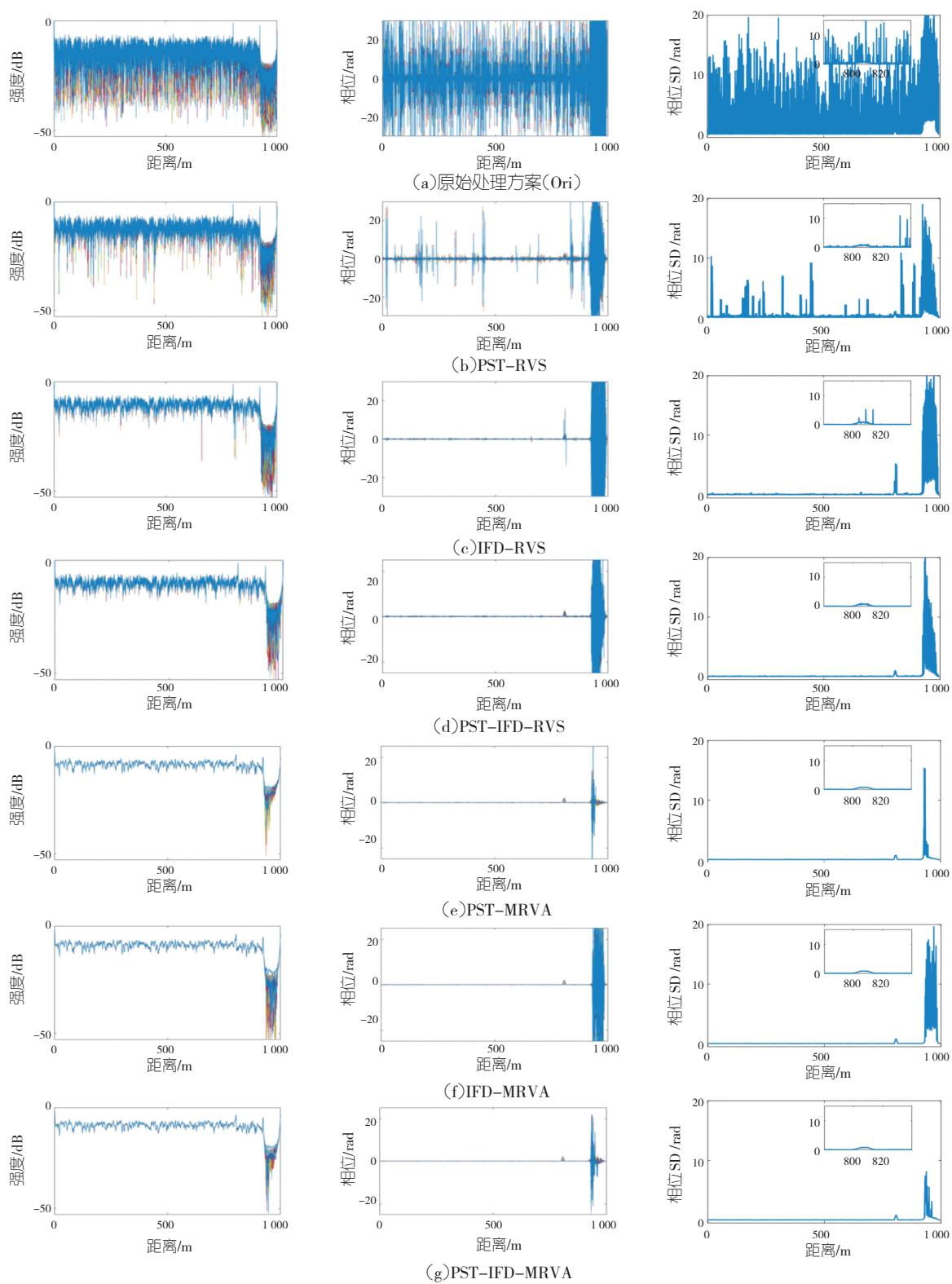


图5 不同衰落抑制方案解调出的环境扰动以及ASD



IFD-MRVA: 脉冲内频分-移动旋转矢量平均

IFD-RVS: 脉冲内频分-旋转矢量和

PST-IFD-MRVA: 相移变换-脉冲内频分-移动旋转矢量平均

PST-IFD-RVS: 相移变换-脉冲内频分-旋转矢量和

PST-MRVA: 移动旋转矢量平均

PST-RVS: 相移变换-旋转矢量和

SD: 标准差

图6 不同方案得到的强度、差分相位以及相位标准差曲线

振动识别上的有效性。图7给出了性能指标的对比结果。首先,在平均幅度波动方面,原始信号高达81.97 dB,而PST-IFD-MRVA方案显著降低到27.26 dB,降幅超过66%,这展现出最强的衰落抑制能力。在强度SNR方面,原始信号仅22.58 dB,所有衰落抑制方案均实现强度SNR提升,其中PST-IFD-MRVA增益最为显著,达到32.41 dB,提升幅度接近10 dB。在应变灵敏度方面,原始方案为445.3 $\mu\text{e}/\sqrt{\text{Hz}}$,而衰落抑制方案均大幅降低了该值。其中,基于MRVA的方案整体表现最好,尤其是PST-IFD-MRVA,可将灵敏度进一步降低至9.09 $\mu\text{e}/\sqrt{\text{Hz}}$ 。在振动SNR方面,基于MRVA的方案可以将性能提升至40 dB左右,其中PST-IFD-MRVA方案

相比原始信号提升超过17 dB,从而显著增强了对微弱振动的识别能力。

4 结束语

本文针对DSM-ISAC中因干涉衰落效应导致传感性能下降与可靠性不足的研究挑战,提出并实验验证了一种接收端多域集成数字信号处理衰落抑制方案。通过依次引入PST、IFD和MRVA,实现干涉衰落的协同抑制。实验结果表明,受益于DSM信号频谱分配的灵活性,通过优化子载波间保护间隔,在920 m标准单模光纤上,实现了36 GBaud DP-16QAM通信信号与线性调频传感脉冲的共波长信道通感融

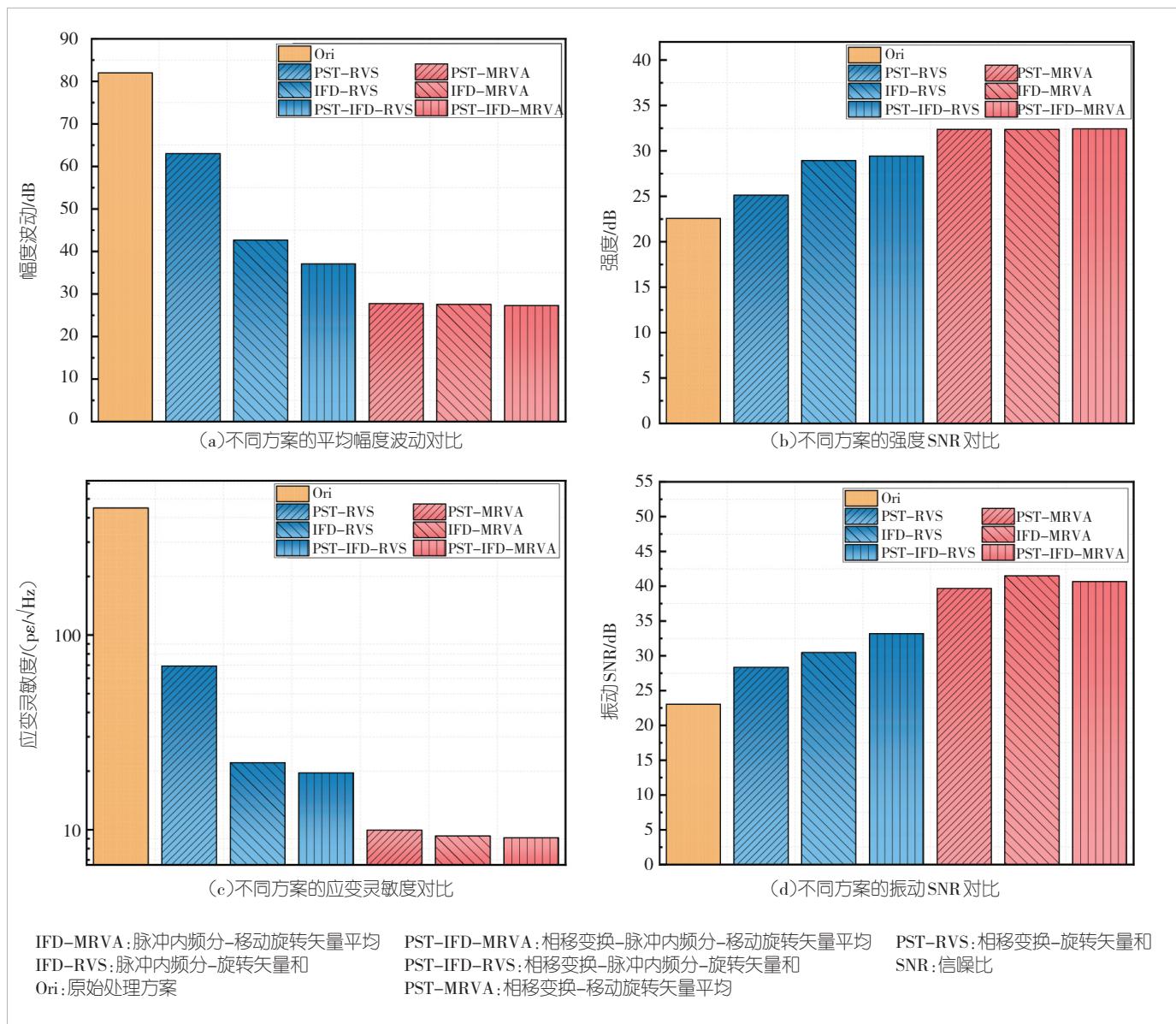


图7 不同衰落抑制方案的性能对比

合。与传统信号的处理方法相比,所提的多域集成衰落抑制方案使信号强度信噪比提升近10 dB,应变灵敏度提高至9.09 $\text{pe}/\sqrt{\text{Hz}}$,并成功解调出10 kHz振动信号,从而显著改善了ISAC系统性能。

参考文献

- [1] LONG J, JIANG L, XIAO C, et al. Large-scale distributed fiber-optic hydrophone towing array and lake trial performance test [J]. *Acta optica sinica*, 2024, 44(1): 0106019
- [2] IP E, HUANG Y K, WELLBROCK G, et al. Vibration detection and localization using modified digital coherent telecom transponders [J]. *Journal of lightwave technology*, 2022, 40(5): 1472–1482. DOI: 10.1109/JLT.2021.3137768
- [3] MAZUR M, CASTELLANOS J C, RYF R, et al. Transoceanic phase and polarization fiber sensing using real-time coherent transceiver [EB/OL]. [2025-11-13]. <https://ieeexplore.ieee.org/abstract/document/9748741>
- [4] IP E, HUANG Y K, WANG T, et al. Distributed acoustic sensing for datacenter optical interconnects using self-homodyne coherent detection [EB/OL]. [2025-11-13]. <https://ieeexplore.ieee.org/abstract/document/9748716>
- [5] GU L, XIANG M, ZHANG Z B, et al. Co-wavelength-channel integration of ultra-low-frequency distributed acoustic sensing and high-capacity communication [EB/OL]. [2025-11-10]. <https://ieeexplore.ieee.org/document/10810040>
- [6] 向梦, 张泽彬, 古龙, 等. 面向甚低频分布式声波感知的标准单模光纤通感融合系统(特邀)[J]. 光学学报, 2025, 45(13): 1306025
- [7] LU Y, ZHU T, CHEN L, et al. Distributed vibration sensor based on coherent detection of phase-OTDR [J]. *Journal of lightwave technology*, 2010, 28(22): 3243–3249. DOI: 10.1109/JLT.2010.2078798
- [8] REN M Q, LU P, CHEN L, et al. Theoretical and experimental analysis of Φ -OTDR based on polarization diversity detection [EB/OL]. [2025-11-09]. <https://ieeexplore.ieee.org/document/7343742>
- [9] ZHOU J, PAN Z Q, YE Q, et al. Characteristics and explanations of interference fading of a Φ -OTDR with a multi-frequency source [J]. *Journal of lightwave technology*, 2013, 31(17): 2947–2954. DOI: 10.1109/JLT.2013.2275179
- [10] XIANG M, YAN S, ZHOU G, et al. EEPN-free carrier phase recovery for spectral-efficient digital subcarrier multiplexing transmissions [J]. *Journal of lightwave technology*, 2025, 43(2): 513–521. DOI: 10.1109/JLT.2024.3466173
- [11] LIU Q, FAN X, HE Z. Time-gated digital optical frequency domain reflectometry with 1.6-m spatial resolution over entire 110-km range [J]. *Optics express*, 2015, 23(20): 25988–25995. DOI: 10.1364/OE.23.025988
- [12] ZHOU J, PAN Z Q, YE Q, et al. Phase demodulation technology using a multi-frequency source for discrimination of interference-fading induced false alarms in a Φ -OTDR system [J]. *Chinese journal of lasers*, 2013, 40(9): 905003. DOI: 10.3788/cjl201340.0905003
- [13] LI H, LIU T, FAN C, et al. Fading suppression for distributed acoustic sensing assisted with dual-laser system and differential-vector-sum algorithm [J]. *IEEE sensors journal*, 2022, 22(10): 9417–9425. DOI: 10.1109/JSEN.2022.3164695
- [14] WANG X, LU B, WANG Z, et al. Interference-fading-free Φ -OTDR based on differential phase shift pulsing technology [J]. *IEEE photonics technology letters*, 2018, 31(1): 39–42. DOI: 10.1109/LPT.2018.2881757
- [15] CHEN D, LIU Q W, HE Z Y. Phase-detection distributed fiber-optic vibration sensor without fading-noise based on time-gated digital OFDR [J]. *Optics express*, 2017, 25(7): 8315–8325
- [16] HE H, YAN L, QIAN H, et al. Suppression of the interference fading in phase-sensitive OTDR with phase-shift transform [J]. *Journal of lightwave technology*, 2020, 39(1): 295–302. DOI: 10.1109/JLT.2020.3023699
- [17] QIAN H, LUO B, HE H, et al. Fading-free Φ -OTDR evaluation based on the statistical analysis of phase hopping [EB/OL]. [2025-11-05]. <https://opg.optica.org/ao/abstract.cfm?uri=ao-61-23-6729>
- [18] YU F, LIU S, XU D, et al. Integrated fading suppression for phase-sensitive OTDR: an algorithm set [J]. *Journal of lightwave technology*, 2024, 42(5): 1754–1764. DOI: 10.1109/JLT.2023.3323394

作者简介



张兵兵, 广东工业大学在读硕士研究生; 主要研究方向为光纤通感融合。



向梦, 广东工业大学教授; 主要从事基于新波段、新光纤、新放大的高速光传输关键技术研究。



张泽彬, 广东工业大学在读硕士研究生; 主要研究方向为光纤通感融合。



付松年, 广东工业大学教授、博士生导师, 中国光学工程学会光通信与信息网络技术专委会常务委员, 《中国激光》《Optics Express》和《IEEE Photonics Journal》编委, 广东省“珠江人才计划”引进创新创业团队负责人; 长期从事光纤通信关键技术研究; 主持“十四五”重点研发计划项目、国家自然科学基金杰出青年科学基金和重点项目。



秦玉文, 广东工业大学教授、博士生导师, 广东工业大学先进光子技术研究院院长, 通感融合光子技术教育部重点实验室主任, 广东省信息光子技术重点实验室主任; 主持国家重点研发计划、国家自然科学基金重点项目, 以及广东省重大专项等项目; 发表论文150余篇, 获授权发明专利60余项。

新一代光传输网络关键器件和集成化技术研究



Key Devices and Integration Technology of Next-Generation Optical Transmission Network

罗勇/LUO Yong, 张冀/ZHANG Ji, 梁雪瑞/LIANG Xuerui

(武汉光迅科技股份有限公司, 中国 武汉 430205)
(Accelink Technology Co., Ltd., Wuhan 430205, China)

DOI: 10.12142/ZTETJ.202506006

网络出版地址: <https://link.cnki.net/urlid/34.1228.TN.20251218.1430.004>

网络出版日期: 2025-12-19

收稿日期: 2025-10-20

摘要: 新一代光传输网络向着超大容量、超长距离、高灵活性和低能耗的方向发展, 其核心驱动力在于关键器件的创新与集成化技术的突破。围绕新一代光传输网络, 系统地综述了光电集成技术平台的特性与选择, 并深入探讨了可调谐光源、集成化收发芯片、集成光波导放大器(EDWA)及智能光交换器件等关键器件的最新研究进展。在此基础上, 分析了这些技术在超大容量传输系统、数据中心互联(DCI)和5G承载网中的具体应用。最后, 总结了当前面临的主要挑战, 如多材料异构集成、热管理和封装成本等, 并展望了共封装光学(CPO)、晶圆级制造及智能化光网络等未来发展趋势。

关键词: 新一代光传输网络; 光电集成; 硅光子; 可调谐激光器; 光收发芯片; 可重构型光分插复用器

Abstract: Next-generation optical transmission networks are evolving toward ultra-high capacity, ultra-long distance, high flexibility, and low energy consumption. These advancements are driven mainly by innovations in key devices and breakthroughs in integration technologies. The characteristics and selection of optoelectronic integration platforms are systematically reviewed, and the latest research progress on key devices, including tunable light sources, integrated transceiver chips, integrated optical waveguide amplifiers (EDWAs), and intelligent optical switching devices, is thoroughly discussed and examined. Based on this analysis, the applications of these technologies in ultra-high-capacity transmission systems, data center interconnections (DCI), and 5G bearer networks are examined. Finally, the major challenges currently faced are summarized such as multi-material heterogeneous integration, thermal management, and packaging costs, and future trends, including co-packaged optics (CPO), wafer-level manufacturing, and intelligent optical networks are prospected.

Keywords: next-generation optical transmission network; optoelectronic integration; silicon photonics; tunable laser; optical transceiver chip; reconfigurable optical add-drop multiplexer

引用格式: 罗勇, 张冀, 梁雪瑞. 新一代光传输网络关键器件和集成化技术研究 [J]. 中兴通讯技术, 2025, 31(6): 39-48. DOI: 10.12142/ZTETJ.202506006

Citation: LUO Y, ZHANG J, LIANG X R. Key devices and integration technology of next-generation optical transmission network [J]. ZTE technology journal, 2025, 31(6): 39-48. DOI: 10.12142/ZTETJ.202506006

随着互联网、物联网、大数据、云计算等新兴技术的飞速发展, 全球数据流量呈现爆发式增长。面对庞大的数据传输需求, 传统光传输网络逐渐暴露出容量有限、传输距离不足、能耗较高等问题。因此, 研发新一代光传输网络已成为全球通信领域的核心任务之一。关键器件作为光传输网络的核心组成部分, 其性能直接决定着网络的传输效率与质量。集成化技术则是解决网络复杂度高、成本居高不下、能耗偏大等一系列问题的重要途径, 能够显著提升系统的可靠性、稳定性与可扩展性。综上所述, 深入地研究新一代光传输网络中的关键器件与集成化技术, 对于推动全球通信网

络的升级、满足日益增长的数据传输需求具有非常重要的现实意义。

1 光网络关键技术发展

1.1 发展历程

自 20 世纪 70 年代第一代光传输网络诞生以来, 光传输技术已历经多次重大变革。从早期基于石英光纤与简单调制格式的低速传输系统, 到随后采用密集波分复用 (DWDM) 和外差检测技术实现长距离、大容量传输的第二代光传输网

络，再到如今融合了多种先进调制解调技术、数字信号处理及智能管控技术的新一代光传输网络，每一次技术演进均伴随着关键器件的创新突破与集成化水平的显著提升。

1.2 关键特性

新一代光传输网络具备以下主要特征：

- 1) 超大容量：系统单通道可实现Tbit/s级乃至更高传输速率，总容量可达数百Tbit/s，可满足全球数据流量的爆发式增长需求。
- 2) 超长距离传输：借助先进的光放大与色散补偿技术，实现跨洲际的超长距离无中继传输，从而降低网络建设和维护成本。
- 3) 高灵活性与可扩展性：支持灵活的网络拓扑结构与动态带宽分配，能够依据业务需求快速调整网络资源配置，便于网络升级与扩展。
- 4) 低能耗：相较于传统光传输网络，在单位比特能耗方面实现大幅降低，符合全球节能减排的发展趋势。

光传输网络呈现更高速率、宽带化、全光化、新型网络融合一体化等趋势，其技术的发展驱动光器件和光芯片集成技术加速更迭。

2 集成化光芯片技术

2.1 平台选择与比较

在光电集成技术中，根据不同材料实现的器件可分为多种平台，以适应多样化功能需求。平台的选择对器件性能、制造成本及应用领域具有重要影响。其中，磷化铟（InP）基光电集成平台具有高效的光电转换特性，尤其适用于高速、高功率光电子器件，其制造工艺成熟可靠，在光通信领域应用广泛。然而，InP平台的集成度相对较低，且成本较高、工艺复杂，制约了其大规模应用。

相比之下，硅光（SiPh）集成平台与现有互补金属氧化物半导体（CMOS）工艺兼容，较高的芯包层折射率差可实现纳米级结构尺寸，有利于实现大规模集成，适用于低成本消费类电子产品。但受限于硅材料的晶格与能带特性，其光电转换效率较低，在下一代高速演进中存在带宽约束问题。

氮化硅（ Si_xN_y ）材料具有较低的光传播损耗，且在高温高功率环境下稳定性良好，适合高频、低损耗光学波导应用。尽管在集成度与性能优化方面不及硅光平台，氮化硅在某些高频、低损耗光通信系统中仍具备显著优势。

二氧化硅（ SiO_2 ）平台以低损耗、机械稳定和温度不敏感等特点著称，尤其适用于长距离光通信。多种商用无源器

件如平面光波导光分路器（PLC Splitter）和阵列波导光栅（AWG）均基于该平台实现。然而， SiO_2 缺乏电光调制能力，且难以实现高密度集成。

薄膜铌酸锂（TFLN）材料具有突出的非线性效应和高电光调制效率，在高速光通信与量子信息处理等高端应用中表现优异。但其制备与CMOS工艺不完全兼容，集成度低，未来可通过与硅光进行异质集成，提升功能与性能。

综上所述，不同光电集成平台各具优势与局限，在实际研发中需根据具体应用场景、性能指标及成本等因素综合选择。当前，多材料融合集成已成为研究热点，通过材料协同与工艺创新突破单一材料限制，有望为下一代信息传输与处理提供关键支撑。表1总结了各平台的优缺点、应用场景及制备技术。

2.2 关键技术

关键技术包括以下6个方面：

- 1) 光无源结构设计技术，采用先进的模拟软件和优化算法，能够实现片上损耗低、光纤耦合效率高的光无源结构。
- 2) 高带宽调制技术，基于新型材料或结构，实现小型化、低偏压、低插损、高带宽光电调制器阵列芯片。
- 3) 高带宽探测技术，设计新型的探测器结构，在保证暗电流和响应度的同时，合理设计本征区，降低载流子渡越时间实现较大光电转换带宽。
- 4) 光电合封技术，将光芯片与电子芯片（如ASIC、交换芯片）集成在同一基板上，通过缩短电信号传输路径来提升性能。其核心目标是通过光电协同设计，解决高速数据传输中的功耗、延迟和带宽瓶颈问题。
- 5) 耦合技术，通过精确的对准和键合工艺，可以实现芯片与光纤、芯片与芯片之间的高效耦合。
- 6) 热管理技术，研发了新型的散热结构和材料，如微通道冷却技术、石墨烯基散热涂层等，有效解决了光电子集成芯片在高速运行时产生的热量问题，保证了器件的稳定性和可靠性。

3 新一代光传输网络关键器件

图1展示了光传输网络核心光器件的技术演进趋势。在集成可调激光器组件（ITLA）方面，随着相干光模块速率从100 Gbit/s提升至800 Gbit/s，激光器线宽从300 kHz压缩到50 kHz，模块封装尺寸亦不断缩小——从C封装可插拔（CFP）发展到双密度四通道小型可插拔（QSFP-DD），对ITLA的出光功率与尺寸提出了更高要求，从而形成了Nano-

表1 光芯片平台特性对比

平台	制备技术	应用场景	优点	缺点
InP	外延生长、光刻、刻蚀、键合集成	相干收发芯片、长距离光传输系统、高速光通信、光放大器	直接带隙,支持激光器、探测器等有源器件,高光电转换效率,适合高速、大带宽通信	材料和工艺成本高,集成难度高
SiPh	CMOS兼容工艺(SOI衬底)、光刻、干法刻蚀、金属互连	数据中心间互联、短距光互连、片上集成光子电路	CMOS工艺兼容,可大规模低成本制造,集成度高,适合大规模光子电路	下一代演进调制带宽受限
Si _x N _y	薄膜沉积、CMP、光刻刻蚀	光学滤波器、微环谐振腔、光频梳、光量子器件、低损耗传输	极低的光传输损耗,宽带透明窗口(400~2 300 nm),良好的温度稳定性,非优秀的线性光学性能	无有源器件功能(需与InP/硅光集成),集成度有限
SiO ₂	等离子体增强化学气相沉积、火焰沉积、离子交换	长距离光通信、光纤型无源器件、PLC分路器(Splitter)、AWG波分复用器	超低光损耗(dB/m级),工艺成熟,热稳定性好,成本低	无电光效应,无法实现有源器件,集成度低,尺寸较大
TFLN	晶体切割、薄膜键合、离子切片、刻蚀(ICP-RIE)、异质集成	高速调制器、微波光子学、量子信息处理、非线性频率转换	极高电光系数(r33~30 pm/V),宽带、低损耗,高速相位/强度调制能力,非线性效应强	成本高,工艺复杂,片上光源和探测器需异质集成,尚未与CMOS工艺完全兼容
AWG:阵列波导光栅 CMOS:互补金属氧化物半导体 CMP:化学机械抛光		ICP-RIE:电感耦合等离子体-反应离子刻蚀 InP:磷化铟 PLC:平面光波导	SiO ₂ :二氧化硅 SiPh:硅光 Si _x N _y :氮化硅	SOI:绝缘衬底上硅 TFLN:薄膜铌酸锂

ITLA 与 Pico-ITLA 的规范要求。

针对相干收发光组件,当前400G光传输网络核心光器件已实现规模化应用,支撑着现有400G DWDM光传输网络。未来其发展将沿两个方向推进:一是速率由128 Gbaud提升至256 Gbaud;二是波长范围由C波段扩展至C+L波段。

3.1 可调谐光源器件

新一代光传输网络要求激光光源具备低噪声、超宽调节范围、更窄谱线宽度及更低功耗。为克服传统半导体光源线宽的问题,高Q值无源外腔反馈被引入并证实为压缩线宽的有效方法。基于外腔光反馈技术的无源外腔可分为分立外腔、组合外腔和集成外腔3类。尽管采用上述非集成腔结构

可实现极窄线宽激光输出,但仍面临尺寸较大、对机械振动敏感、耦合损耗高、成本昂贵、调节范围有限、功耗偏大等问题。

相较之下,集成外腔反馈激光器在实现窄线宽的同时,显著减小了外腔的尺寸、体积与重量。硅光集成外腔可通过半导体批量工艺在单一衬底上制备,具备与其他光电子组件集成的能力,从而提升了系统的稳定性与可靠性。因此,硅光集成外腔可调激光器已成为下一代可调谐光源的关键技术。

从工艺实现角度,如图2(b)与(c)所示,硅基集成外腔可调激光器可分为基于对接耦合技术的混合集成激光器和基于外延生长的异质集成激光器。其中,对接耦合技术发展较为成熟,具备对增益芯片和外腔分别进

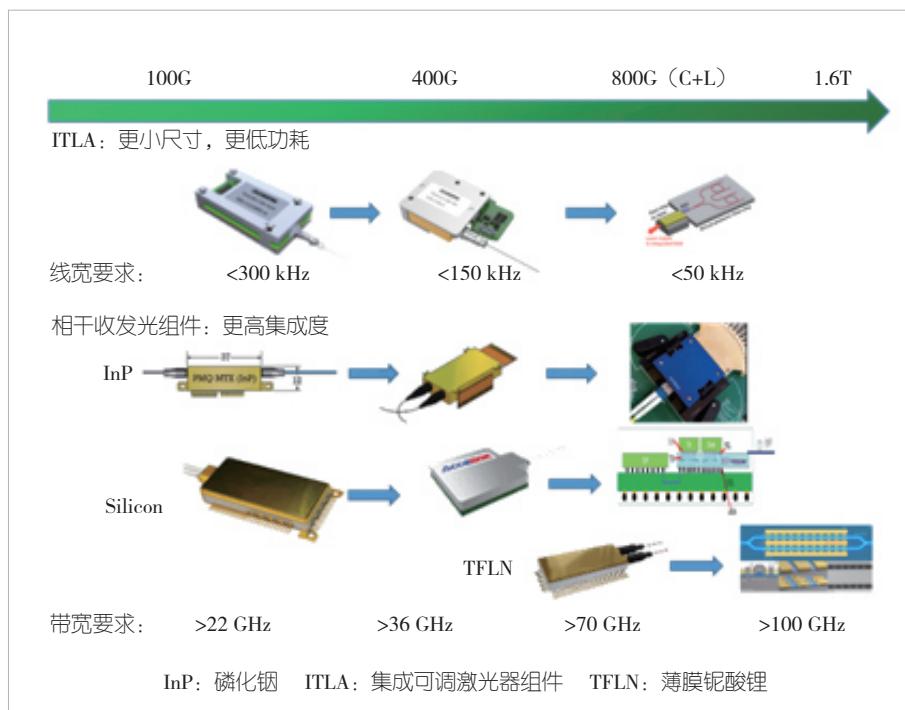


图1 光传输网络核心光器件技术演进趋势

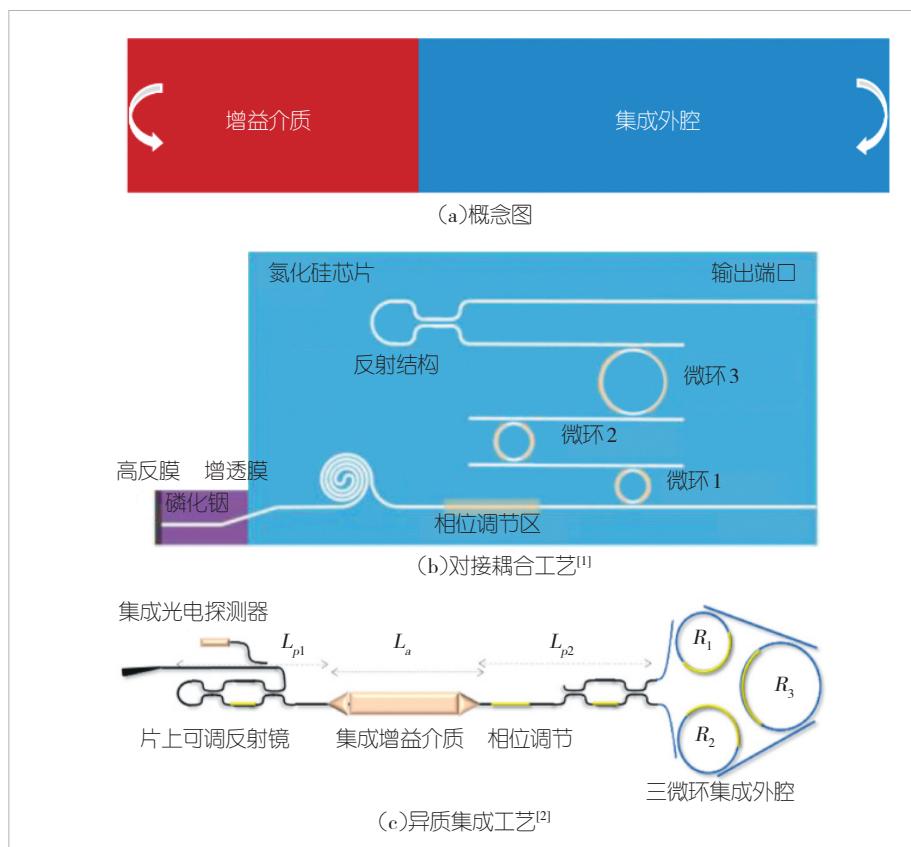


图2 集成外腔反馈半导体激光器及典型集成外腔结构

行优化的灵活性，因而备受关注。目前，硅、 SiO_2 和 Si_3N_4 等多种介质波导平台已被用于制备外腔，形成混合集成激光器^[1]。该工艺的关键难点在于尽可能降低增益芯片与外腔芯片间的对接耦合损耗，以确保激光器整体性能。在此方向开展代表性研究的有NEC公司、东北大学、清华大学、康奈尔大学及加州大学伯克利分校等。

异质集成技术通过将未加工的III-V族材料直接键合于硅波导顶部，或在硅晶圆上直接外延生长III-V族材料，实现与硅基CMOS工艺兼容，从而有效降低制造成本。与混合集成相比，在保证键合质量的前提下，异质集成器件对冲击、振动等具有更好的鲁棒性。

然而，键合质量实际上受诸多因素影响，如键合工艺过程、晶格失配、极化失配以及材料热膨胀系数差异等。与对接耦合技术相比，目前尚未有基于异质集成技术的商用激光器问世。如表2所示，异质集成已成为当前学术研究的热点，其有望通过倏逝波耦合技术降低耦合损耗与反射，进而提升激光器线宽与出光功率。随着键合技术的持续进步，若未来能实现可控的工艺良率，异质集成技术有望广泛应用于实际产品中。

3.2 集成化收发芯片与器件

随着云计算与人工智能的快速发展，超大规模数据中心集群的数据流量已提升至P比特量级，对超大容量、低功耗光互联的需求日益迫切。集成光子技术凭借其高带宽、低损耗等特性，成为应对上述需求的核心技术路径。集成化光收发芯片与器件主要基于InP、硅及铌酸锂等材料平台实现，以下简要介绍基于这3种材料的集成化光收发芯片与器件的研究进展。

InP光子集成技术具备高光电转换效率、大带宽与低功耗等优势，是支撑AI数据中心算力持续增长的关键推动力之一。日本NTT研究所于2023年研制出基于InP的相干驱动调制器与相干接收器，如图3 (a) 所示，该器件3 dB带宽超过了80 GHz，并实现了128 Gbaud的16QAM调制，刷新了当时同类器件的带宽记录^[8]。2024年，德国弗劳恩霍夫协会提出一种具有平坦频率响应的128 GBaud相干接收机，该器件采用InP光波导集成相干光电探测器与双线性SiGe跨阻放大器，带宽同样超过80 GHz^[9]，如图3 (b) 所示。此外，美国Lumentum公司在2025年OFC展会上展示了单通道400 Gbit/s光链路及光学共封装器件，显著提升了光通信系统的传输速率与稳定性，适用于高速大容量传输场景。

基于硅光平台的收发集成芯片能够单片集成高速调制器与探测器，并借助2D/3D先进光电封装技术实现高密度、高集成度的光收发器件，该方案已成为当前光电器件的主流解

表2 集成硅基外腔激光器性能

作者	工艺技术	边模抑制比/dB	调节范围/nm	最小线宽/kHz
ZHU ^[3]	混合集成	52	46	18
ZHU ^[4]	混合集成	50	47	85
FAN ^[1]	混合集成	60	70	<0.04
LIANG ^[5]	异质集成	>40	40	150
TRAN ^[6]	异质集成	>50	55	50
HUANG ^[7]	异质集成	>55	—	1
TRAN ^[2]	异质集成	>45	40	2
	异质集成	>40	110	<0.22

解决方案之一。硅光集成芯片具备除光源外的全功能集成能力，可包含马赫-曾德尔（MZ）硅光调制器、微环调制器、锗探测器（Ge PD）、多模干涉耦合器、热光调制器、硅波导、氮化硅波导及波导转换结构、端面耦合器、光栅耦合器等各类有源与无源单元。图4展示了台积电硅光平台所集成的典型功能单元^[10]。基于硅光芯片的光电共封装技术，主要采用陶瓷基板、有机基板、玻璃基板或硅基板将电芯片与硅光芯片进行系统集成。封装互连技术也从传统金丝键合向球栅阵列（BGA）等更高密度的方式演进，封装架构正由二维向三维光电共封装集成方向发展。

图5为光迅公司开发的、基于硅光技术的400 Gbit/s与800 Gbit/s相干光收发集成器件。该器件内部集成硅光相干收发芯片及高速驱动芯片（Driver）、跨阻放大器（TIA）电芯片，采用2D倒装先进光电共封装工艺，实现BGA形态的封装。作为数据中心间高速光互连的核心光器件，该器件主要应用于400ZR/ZR+和800ZR相干光模块，支持在400 Gbit/s 16QAM调制格式下实现长达450 km的传输距离，以及在800 Gbit/s 16QAM调制格式下实现长达120 km的传输。

美国Acacia公司采用先进的三维堆叠封装技术，将相干高速光电组件全部集成于单一封装内，如图6所示。该结构通过将数字信号处理器（DSP）裸片、光子集成芯片（PIC）、驱动器及TIA进行一体封装，显著缩小了光器件尺寸。同时，三维堆叠设计能够有效缩短射频（RF）接口之间的互连长度，从而改善其在130 GBaud等高波特率下的信号完整性。

与InP和硅光平台相比，薄膜铌酸锂（LiNbO₃）凭借其高电光系数、强模场约束能力以及宽透明波长窗口，已经成为高性能光子集成的重要平台，这就是下一代200 GBaud应用的关键调制技术。2023年，贝尔实验室研究团队基于TFLN平台的IQ调制器实现了260 GBaud单波长相干传输（调制器实物如图3（c）所示），并在100 km标准单模光纤

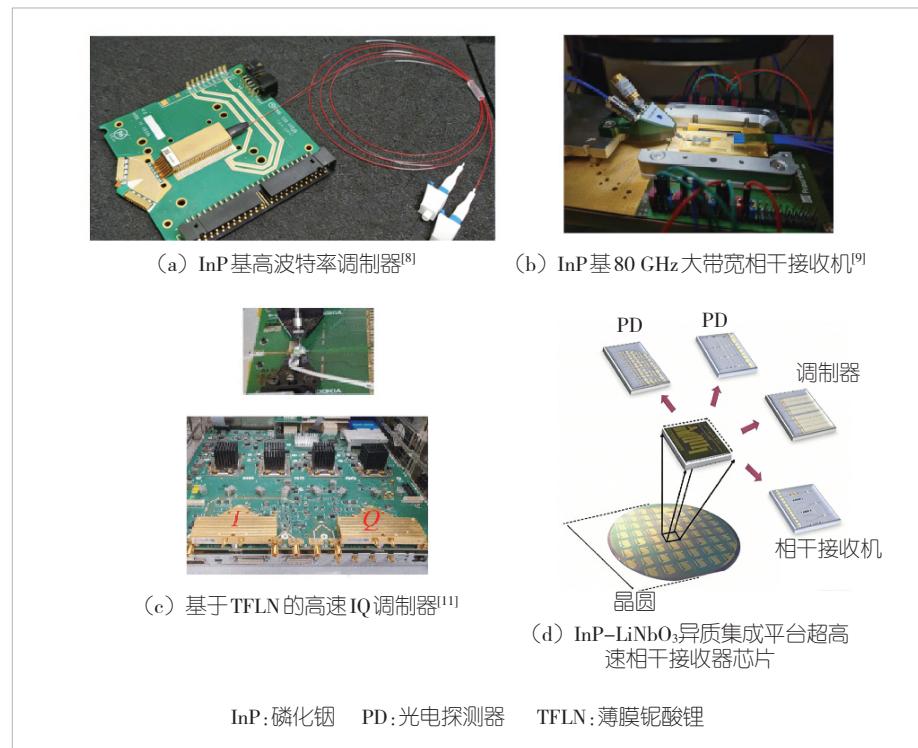


图3 InP材料和TFLN材料集成化光收发芯片结构图

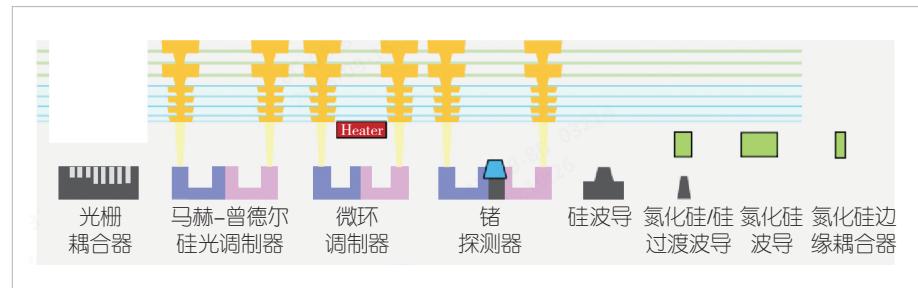


图4 TSMC硅光平台集成化功能单元

上实现了800 Gbit/s的净传输速率。该工作首次在不依赖复杂数字信号处理的情况下，创下260 GBaud的相干传输记录^[11]。

2025年，以色列PacketLight Networks与日本富士通公司合作共同提出了一种基于薄膜铌酸锂与板载光学技术的800 Gbit/s光收发器件，并在光纤通信会议及展览会（OFC）上进行了演示。同年，中国铌奥光电在OFC展会上发布了基于TFLN平台的最新3.2 Tbit/s直调直检光收发器件。西南交通大学于2025年在InP-LiNbO₃混合集成平台上，单芯片集成了超高速相干接收器芯片，突破了TFLN平台在高速有源器件集成方面的核心瓶颈，实现了140 GHz超宽带光电探测器（PD）与3.584 Tbit/s相干接收芯片（器件结构示意如图3（d）所示）。相比基于InP、SiGe等平台的相干接收芯

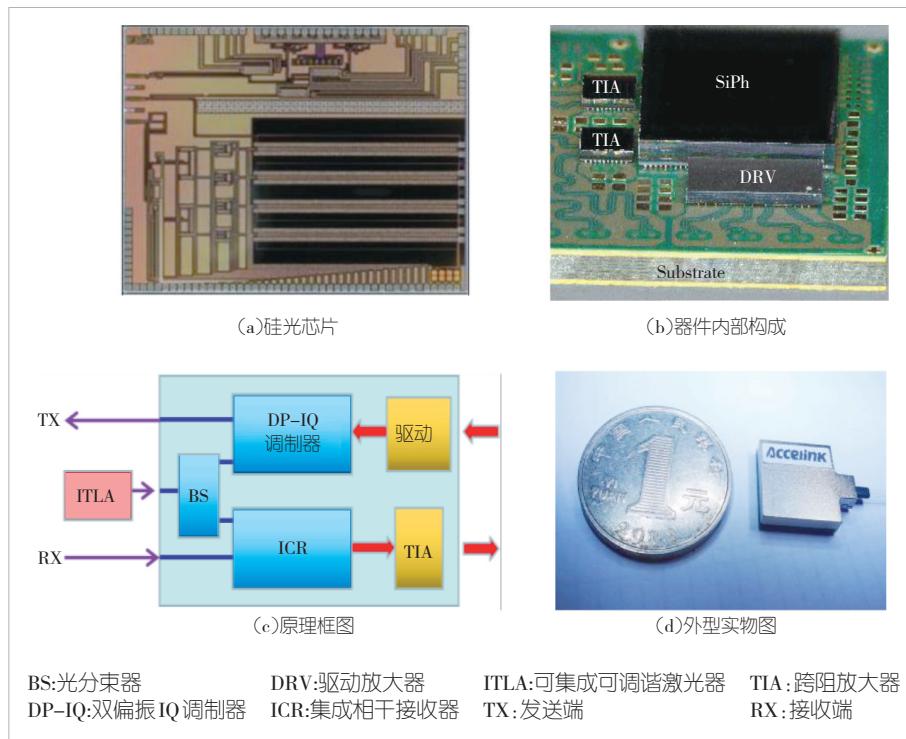


图5 基于2D共封装的硅光相干光电集成器件

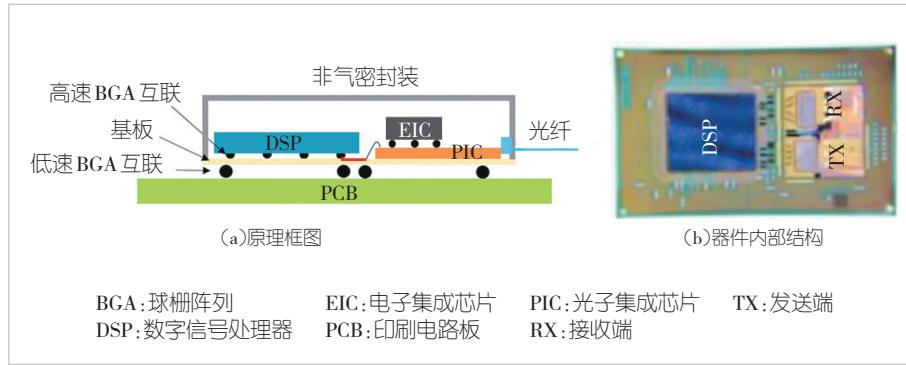


图6 基于3D共封装的硅光相干光电集成器件

片,该异质集成方案在关键指标上取得了一些显著突破:单PD带宽提升75% (达140 GHz),单通道容量(600 Gbit/s)与总容量(3.584 Tbit/s)均创下纪录,且能耗降低30%以上。其晶圆级制造流程也突破了传统芯片级集成在成本与可扩展性方面的限制,为大规模应用奠定了基础^[12]。

3.3 新型光放大器件

基于传统C波段的掺铒光纤放大器(EDFA)的DWDM系统已难以满足云计算与数据中心等高要求业务对传输容量急剧增长的迫切需求,光放大技术正面临波段扩展与新型光纤连接等多方面挑战。当前在光通信领域中,EDFA光放大器的核心部件——泵浦激光器仍高度依赖进口,因此研发高

功率、大模场单模泵浦激光器芯片及封装技术是降低系统成本的关键因素。随着波段扩展及L波段EDFA的尺寸限制,超宽谱低噪声半导体光放大器(SOA)芯片正逐步走向系统商用。例如,华为设计了一款基于SOA阵列的双端口C+L双偏振放大器模块,具备低损耗、低光串扰和无热退化等特性,但在多波长放大时仍会面临由SOA非线性引起的光信噪比(OSNR)代价问题^[13]。此外,掺稀土元素光波导放大器因具备易于集成、体积紧凑等优势,随着流片工艺水平的不断提升,正逐渐成为当前的研究热点。

集成光波导放大器(EDWA)是一种基于EDFA的集成光学器件,适用于需要高增益与长距离传输的小型化场景,尤其能够应对L波段EDFA在集成度方面的挑战。EDWA通过片上集成技术提升了光放大器的性能,具备更紧凑的尺寸与更低的功耗。随着光通信系统对带宽与传输速率需求的持续增长,EDWA已成为集成光学器件研究的重要方向之一,有望填补集成光子芯片在低噪声光放大技术方面的空白。

在过去30年里,多家科研团队持续致力于稀土离子掺杂波导放大器的研究。20世纪90年代,美国贝尔实验室率先开展了关于EDWA的工作。然而,受限于当时采用的基于低折射率SiO₂的材料体系,其器件体积大、损耗高,且无法与现代微纳光子集成工艺兼容,相关研究随后陷入停滞。

2022年,瑞士洛桑联邦理工学院在0.5 m长的超低损耗氮化硅集成光波导中,采用了高能铒离子注入实现掺杂,获得了超过145 mW的片上连续光输出功率,并在毫米尺度上实现了高于30 dB的片上连续光增益^[14]。2025年,华东师范大学团队首次在掺铒铌酸锂薄膜光波导放大器中实现了高外部净增益(>18 dB)与低噪声系数(5 dB)。

尽管EDWA展现出可观的应用潜力,其在实际应用中仍面临诸多挑战。首先,不同材料波导中铒离子掺杂的均匀性与浓度控制是难点,掺杂分布不均会导致增益起伏与性能

不稳定。其次，随着集成度提升，EDWA的功耗与热效应问题日益突出。此外，由于波导双折射效应、材料增益固有各向异性，以及模场分布与有源区重叠的优化存在矛盾，EDWA的偏振相关增益（PDG）难以在整个工作带宽内维持在较低水平。因此，如何在高增益、低PDG、低功耗和良好热稳定性之间取得综合优化，仍是当前研究人员面临的重要挑战。

3.4 集成化光无源器件

基于传统SiO₂平台的光无源芯片及器件，以其超低传输损耗、良好的环境可靠性与成熟的低成本工艺，在光传输网络的演进中始终保持重要地位。即使作为功能单一的无源器件，其在系统升级中仍不可或缺。以阵列波导光栅（AWG）为例，随着光网络容量需求的提升，其光谱特性已从传统C波段平顶型向C+L超宽带型演进，通过增加信道数量与提升单信道速率，共同推动网络容量的扩展。

随着新一代光网络对智能化管控要求的提高，多功能集成化正成为光无源器件发展的新趋势。光网络中对于多信道的监控、调度与保护需求，催生了各类集成化光无源器件的发展。

信道监控主要包括信道能量监控、波长监控与偏振态监控；信道的调度与保护则涉及信道能量的控制及光路切换。信道能量监控通常借助特定分光比的分束器（Splitter）或抽头（Tap）功能的无源芯片结构^[15]，并与光电探测器（PD）通过光路耦合与集成封装技术共同实现。

波长监控基于的原理主要包括相干波长检测与相干相位检测两类。前者采用基于180°光混频器结构的相干波长检测技术^[16]；后者则通过对至少两路具有不同相位延迟的马赫-曾德尔干涉仪（MZI）光路进行相位检测与差值定标，再通过检测差值与定标值的比对实现波长判定。为进一步扩大检测范围，已衍生出检测波长范围嵌套型^[17]与叠加型^[18]等技术。

偏振态监控通常借助偏振分束结构^[19]与偏振旋转结构^[20]来实现。

信道的调度与保护主要通过可调光衰减器（VOA）实现强度调控，并通过光开关（OSW）完成光路的切换与保护。

通过将上述具备基础功能的无源光芯片结构单元进行集成，可衍生出多种新型集成化光无源器件形态，为光网络向智能化发展提供关键技术支撑。

几种应用实例包括：

1) 基于VOA、OSW、分束器、抽头多功能集成结构的池化波分信道监控集成化光无源器件产品；

2) 基于VOA、OSW、分束器、抽头、光混频器结构的波长监控与信道保护集成化光无源器件产品；

3) 基于偏振分束器（PBS）、光混频器结构和带有相位延迟线MZI型波长锁定功能结构的光信道监控产品。

以上所述的多功能集成化光无源器件可在同质材料平台内实现。若要进一步提升光无源器件的集成度，并充分发挥不同材料平台功能结构的性能优势、实现功能互补，异质集成技术则成为该领域的核心发展方向。同时，异质集成也有助于突破无源与有源集成之间的技术瓶颈。目前实现异质集成的主要技术路线包括：微转移打印^[21]、晶圆键合、倒装芯片以及直接外延生长等。

3.5 智能光交换与ROADM器件

可重构光分插复用器（ROADM）可实现波长级动态调度，但仍面临多维冲突与阻塞等问题。新一代基于波长选择开关（WSS）的ROADM，结合软件定义网络（SDN）控制器实现光层自动化组网，已在城域网中大规模部署。WSS产品在发展过程中形成了多种技术方案。目前技术成熟且已大规模商用的主要基于空间光调制器的WSS产品，按其调制器类型主要可分为3种：图7（a）基于微机电系统（MEMS）的WSS；图7（b）基于液晶堆叠（LC）的WSS；图7（c）基于硅基液晶（LCOS）的WSS。其中，基于MEMS的WSS因像素尺寸较大，难以适应灵活栅格的光网络需求；基于LC的WSS则受限于液晶单元数量，难以满足网络节点对高端口数量的要求。

相比之下，基于LCOS实现的WSS兼具灵活栅格、多端口支持和高频谱利用率等显著优势，已成为构建下一代弹性光网络的主流选择。未来几年，ROADM设备将主要朝着维度增加与波段扩展等方向发展：维度提升主要指WSS模块端口数量从1×20向1×48或1×64演进；同时，随着中国光通信系统逐步从C波段向L波段扩展，WSS模块的维度与带宽需求也将持续提升。

LCOS作为WSS的核心元件，其芯片性能具体包括分辨率、反射率与刷新率，都对WSS的技术指标具有决定性影响。LCOS结构如图8所示，LCOS分辨率从2K提升至更高，可在实现C+L一体化波段扩展的同时，进一步优化工作带宽。

当前主流的WSS产品主要工作在C波段，而L波段的WSS产品也已实现商业化应用。随着400G时代的成熟和800G技术的逐步演进，C+L波段的WSS产品正逐步进入市场并扩大应用规模。C+L波段WSS的实现难点在于，需在维持信道带宽的条件下扩展工作波长范围。从WSS的基本原

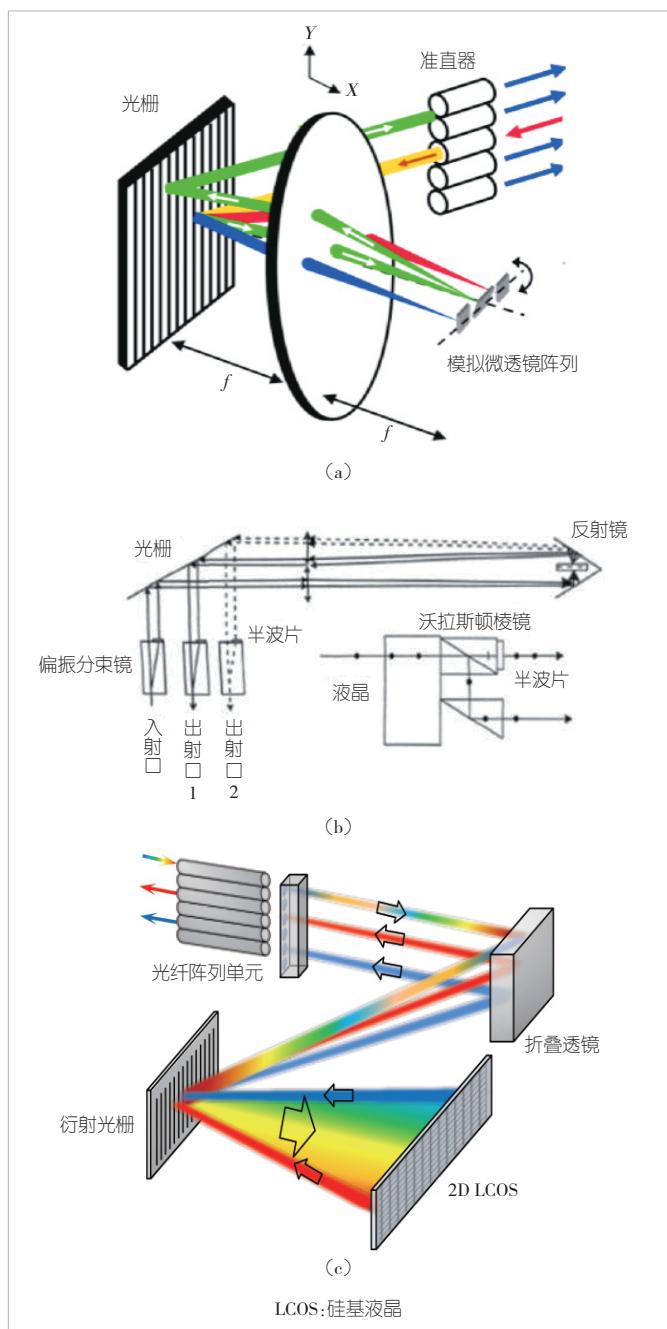


图7 波长选择开关技术方案

理出发,其主要技术手段是增大LCOS芯片的有效使用面积。在此基础上,通过进一步优化光学设计并改进封装工艺,可确保C+L波段WSS在插损、偏振相关损耗及带宽性能等其他关键指标上满足系统要求。图9展示了一款42端口C+L波段WSS的带宽性能。

4 面临的挑战

尽管新一代光传输网络器件与集成技术已取得显著进

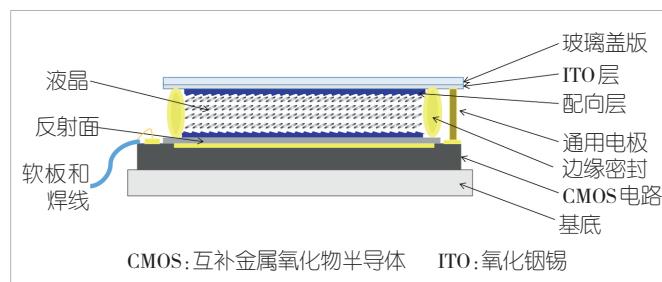


图8 硅基液晶器件像素单元的基本结构示意图

展,但仍面临诸多挑战:

1) 多材料异质集成挑战:将InP、TFLN等多种高性能材料集成到统一的硅基平台上,是发挥各自性能优势的关键路径,但该过程面临晶格失配、热膨胀系数差异、键合工艺复杂及质量控制困难等问题。应对策略包括发展先进的晶圆级键合与选择性区域外延生长技术,优化工艺流程,以提高集成良率和可靠性。

2) 功耗与热管理挑战:随着集成度快速提升,芯片功耗密度显著增大,产生的热量会影响激光器、调制器等核心器件的性能与工作寿命。需通过光电协同设计优化电路与器件结构以降低功耗,同时开发微通道冷却、三维堆叠散热等先进热管理技术,确保芯片稳定运行。

3) 封装与测试挑战:高密度光电芯片中,封装成本占总成本比重较高。光纤阵列与芯片的高效低成本耦合、高速电信号封装以及晶圆级光电测试均构成技术难点。应推动标准化封装流程与接口,发展晶圆级测试技术,引入自动化封装设备,以降低制造成本并提升生产效率。

4) 成本与规模化生产挑战:尽管硅光技术与CMOS工艺兼容,但TFLN、InP等平台的加工成本仍然较高。如何平衡性能、成本与规模化生产之间的关系,是相关技术能否实现广泛商业化应用的关键。需通过持续优化制造工艺、扩大晶圆尺寸、建立标准化工艺设计套件(PDK)等方式,降低设计门槛与流片成本。

5 结束语

5.1 研究工作总结

本文围绕新一代光传输网络中的关键器件与集成化技术展开研究,从光发射器件、光接收器件、光放大器件及光无源器件等关键方向的最新进展入手,系统探讨了光电子集成技术在提升系统容量、降低成本与能耗方面的核心作用。研究表明,新一代光传输网络关键器件与集成化技术的持续发展,为构建高速、高效、可靠的下一代通信网络奠定了坚实

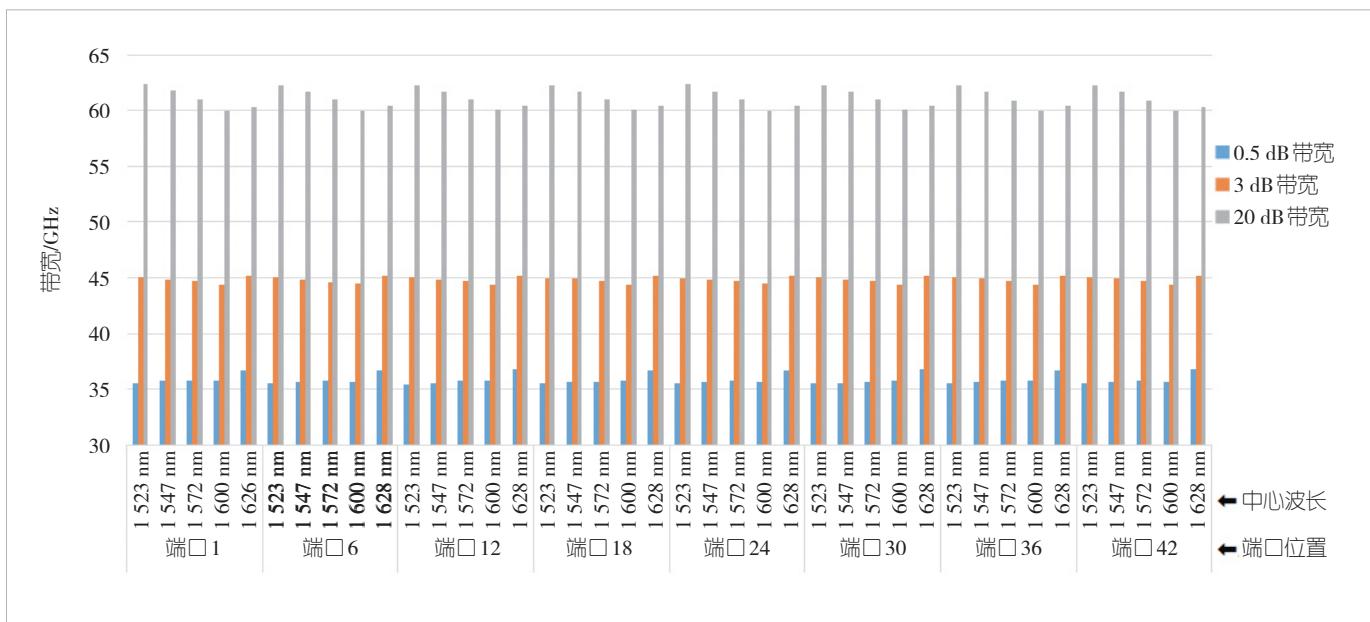


图9 C+L波长选择开关不同工作状态下的原始带宽指标

的技术基础，有力推动了全球通信网络的升级与演进。

驱动力。

5.2 未来发展趋势展望

先进封装技术支撑更高集成度的相干光收发器件：光电共封装是相干光收发器件发展的明确方向，其最终目标是实现光电子处理器的单片集成，从而无缝融合光层功能与数字处理能力。

随着单通道通信速率从400 Gbit/s向800 Gbit/s乃至1.6 Tbit/s迈进，系统不仅需要更优的链路信号完整性设计和更强大的数字信号处理芯片，也对调制器与探测器的带宽性能提出了更高要求。未来，具备极高调制带宽的新材料将获得更广泛的应用前景。其中，铌酸锂（LN）、掺镧锆钛酸铅（PLZT）、钛酸钡（BTO）等薄膜型电光材料将在下一代调制器技术中发挥关键作用，通过异质集成高性能电光材料，硅基光电子器件有望在下一代光收发系统中焕发新的活力。

激光器性能多维度提升：工作波长范围持续拓展，单片集成外腔芯片进一步压缩线宽、缩小尺寸并降低功耗，推动光传输系统在容量与距离上同步提升。

光传输器件波段扩展，逐步实现C+L一体化：包括C+L波段调制与解调技术、C+L波段单片外腔集成可调激光器芯片、适用于超宽带波长选择开关的高分辨率硅基液晶，以及新型光放大技术等。

总之，新一代光传输网络正处于技术革新与产业应用相互促进的关键阶段。持续的技术创新，尤其在多材料异质集成与光电协同设计方面，将成为未来网络演进的核心

致谢

感谢武汉光迅科技股份有限公司王永晨、赵先梦、杨睿、肖礼、陈宏刚、宋梦洋、江毅等专家对本研究工作的支持！

参考文献

- [1] FAN Y, VAN REES A, VAN DER SLOOT P J M, et al. Hybrid integrated InP-Si3N4 diode laser with a 40-Hz intrinsic linewidth [J]. Opt express, 2020, 28(15): 21713 – 21728. DOI: 10.1364/oe.398906
- [2] TRAN M A, HUANG D N, GUO J, et al. Ring-resonator based widely-tunable narrow-linewidth Si/InP integrated lasers [EB/OL]. [2025-11-20]. <https://ieeexplore.ieee.org/abstract/document/8805353>
- [3] ZHU Y Y, ZHU L. Narrow-linewidth, tunable external cavity dual-band diode lasers through InP/GaAs-Si3N4 hybrid integration [J]. Optics express, 2019, 27(3): 2354. DOI: 10.1364/oe.27.002354
- [4] ZHU Y Y, ZENG S W, ZHU L. Optical beam steering by using tunable, narrow-linewidth butt-coupled hybrid lasers in a silicon nitride photonics platform [J]. Photonics research, 2020, 8(3): 375. DOI: 10.1364/prj.382852
- [5] LIANG L J, HULME J, CHAO R L, et al. A direct comparison between heterogeneously integrated widely-tunable ring-based laser designs [EB/OL]. [2025-11-15]. <https://ieeexplore.ieee.org/document/7937313>
- [6] TRAN M A, KOMLJENOVIC T, HUANG D N, et al. A widely-tunable high-SMSR narrow-linewidth laser heterogeneously integrated on silicon [EB/OL]. [2025-11-16]. <https://ieeexplore.ieee.org/document/8426228>
- [7] HUANG D N, TRAN M A, GUO J, et al. High-power sub-kHz linewidth lasers fully integrated on silicon [J]. Optica, 2019, 6(6): 745. DOI: 10.1364/OPTICA.6.000745
- [8] OXAKI J, OGISO Y, HASHIZUME Y, et al. Over-85-GHz-

- bandwidth InP-based coherent driver modulator capable of 1-Tb/s/λ-class operation [J]. Journal of lightwave technology, 2023, 41(11):3290–3296. DOI: 10.1109/JLT.2023.3236962
- [9] GLASEL J, SCHINDLER A, BOERMA H, et al. 128 GBaud coherent receiver engine with flat frequency response [C]//2024 Optical Fiber Communications Conference and Exhibition (OFC). IEEE, 2024. DOI: 10.1364/OFC.2024.W1K.7
- [10] LU H Y, TANG C T, CHEN C H, et al. Low-loss high-uniformity silicon nitride optical building blocks integrated on silicon photonics platform [EB/OL]. [2025-11-10]. <https://opg.optica.org/abstract.cfm?URI=OFC-2025-Th1G.2>
- [11] ALMONACIL S, MARDOYAN H, JORGE F, et al. 260–GBaud single-wavelength coherent transmission over 100-km SSMF based on novel arbitrary waveform generator and thin-film niobate I/Q modulator [J]. Journal of lightwave technology, 2023, 41(12): 3674–3679. DOI: 10.1109/JLT.2023.3269740
- [12] XIE X J, WEI C, HE X C, et al. A 3.584 Tbps coherent receiver chip on InP–LiNbO₃ wafer-level integration platform [J]. Light science applications, 2025, (14): 172. DOI: 10.1038/s41377-025-01821-1
- [13] AZOUIQUI S L, HRAGHI A, DEMIRTZIOGLOU I, et al. Design and integration of a two-port C+L high performance amplifier in a module [EB/OL]. [2025-11-09]. <https://ieeexplore.ieee.org/document/11263156>
- [14] LIU Y, QIU Z R, JI X R, et al. A photonic integrated circuit based erbium-doped amplifier [J]. Science, 2022, 376(6599): 1309–1313. DOI: 10.1126/science.abe2631
- [15] TANG X G, LIAO J K, LI H P, et al. A novel scheme for 1 × N optical power splitter [J]. Optics express, 2010, 18(21): 21697. DOI: 10.1364/oe.18.021697
- [16] KURATA Y, NASU Y, TAMURA M, et al. Heterogeneously integrated PLC with low-loss spot-size converter and newly developed waveplate PBS for DC–DP–16QAM receiver [J]. Journal of lightwave technology, 2015, 33(6): 1202–1209. DOI: 10.1109/JLT.2014.2380814
- [17] IP E, LAU A P T, BARROS D J F, et al. Coherent detection in optical fiber systems [EB/OL]. [2025-11-09]. <https://opg.optica.org/oe/fulltext.cfm?uri=oe-16-2-753>
- [18] STERN B, KIM K, GARIAH H, et al. Athermal silicon photonic wavemeter with wide temperature range [C]//Proceedings of 2023 Optical Fiber Communications Conference and Exhibition (OFC). IEEE, 2023: 1–3. DOI: 10.23919/ofc49934.2023.10116501
- [19] WANG Q C, LU J W, CHEN Y F, et al. High-precision wide-range FBG interrogation based on AWG and ultra-lightweight machine learning model [EB/OL]. [2025-11-15]. <https://www.scencedirect.com/science/article/abs/pii/S0030399225009314?via%3Dihub>
- [20] HASHIZUME Y, KASAHIARA R, SAIDA T, et al. Integrated polarisation beam splitter using waveguide birefringence dependence on waveguide core width [J]. Electronics letters, 2001, 37(25): 1517–1518. DOI: 10.1049/el:20011012
- [21] TAN Y, NIU S P, BILLET M, et al. Micro-transfer printed thin film lithium niobate (TFLN)–on–silicon ring modulator [J]. ACS photonics, 2024, 11(5): 1920–1927. DOI: 10.1021/acsphotonics.3c01869

作者简介



罗勇,武汉光迅科技股份有限公司光电技术研发部总经理,正高级工程师,湖北省通信协会理事;长期从事光电子器件设计及工艺技术开发、器件可靠性与失效分析等工作;多次负责国家“863”计划项目、国家“973”计划项目等;申请专利38项。



张冀,武汉光迅科技股份有限公司光迅丹麦产品线经理,高级工程师;长期从事光波导芯片设计及开发、光电子器件设计及封装工艺技术开发等工作,主导开发了多款无热AWG模块产品、数据通信用CWDM波分芯片、池化波分应用的CS光模块、传感应用光芯片产品;发表论文3篇,申请专利11项。



梁雪瑞,武汉光迅科技股份有限公司器件研发部副经理,高级工程师;长期从事高速硅光芯片、薄膜铌酸锂芯片以及光电集成技术的研发及产品化工作;承担重大科研项目10余项;曾获湖北省科技进步奖一等奖、中央企业优秀团队奖、“3551光谷人才计划”优秀青年人才奖等;申请国家发明专利30余项。



关于AI原生的几点探讨

Several Insights into AI Nativism

何宝宏/HE Baohong

(中国信息通信研究院, 中国北京 100191)
(China Academy of Information and Communications Technology, Beijing 100191, China)

DOI: 10.12142/ZTETJ.202506007

网络出版地址: <https://link.cnki.net/urlid/34.1228.TN.20251219.1010.004>

网络出版日期: 2025-12-19

收稿日期: 2025-10-15

摘要: 人工智能的应用模式主要分为两类: 一是助力传统产业的升级改造, 实现存量系统的智能化; 二是依托智能原生催生出新技术、新产品和新服务, 构建全新的智能系统。智能原生的概念虽已被广泛接受, 但相关探索仍处于相对初级阶段, 其理论内涵还不够丰富。本文从技术、业务和人三个角度, 探讨对智能原生的观察和思考, 并展望其发展趋势, 希望能对AI的应用落地和模式创新起到一定的参考作用。

关键词: 智能原生; 原生多模态; 数据战略; 智能算力; AI人

Abstract: The application modes of artificial intelligence are mainly divided into two categories: one is to facilitate the upgrading and transformation of traditional industries and realize the intelligent transformation of legacy systems; the other is to spawn new technologies, products and services based on intelligent nativism, so as to build brand-new intelligent systems. Although the concept of intelligent nativism has been widely accepted, relevant research is still in a relatively preliminary stage and its theoretical connotation remains insufficiently enriched. From the three perspectives of technology, business and human beings, this paper discusses the observations and reflections on intelligent nativism, looks forward to its development trends, and hopes to provide a certain reference for the application implementation and mode innovation of AI.

Keywords: intelligent nativism; native multimodality; data strategy; intelligent computing power; AI human

引用格式: 何宝宏. 关于AI原生的几点探讨 [J]. 中兴通讯技术, 2025, 31(6): 49–52. DOI: 10.12142/ZTETJ.202506007

Citation: HE B H. Several insights into AI nativism [J]. ZTE technology journal, 2025, 31(6): 49–52. DOI: 10.12142/ZTETJ.202506007

数字化转型旨在通过互联网、云计算、大数据和物联网等信息技术的应用, 对企业等主体的业务流程、商业模式和运营管理进行全面重构, 以提高效率、降低成本和提升竞争力。这一全球性的浪潮已经持续了20余年, 未来几年仍将保持高速增长的态势^[1]。

数字化转型是数字技术对已有系统的赋能, 而数字原生是相对于数字化转型而言的, 它更强调土生土长于数字时代并以数字技术为基础或核心的全新系统。数字原生(代)最初指那些从小就生活在数字化环境中, 天然地对计算机、智能手机和互联网等数字产品有亲近感的一代人。但经过20余年的发展演进, 数字原生的概念已经从专指一代人扩展到技术、业务和生态等数字世界的方方面面, 泛指那些从诞生之初, 就基于数字技术或以数字技术为核心的新业态^[2]。

近年来, 随着以大模型为代表的生成式人工智能的发展成熟, 传统的数字化转型和数字原生的概念, 已难以适配人工智能(AI)时代的新变化与新需求, 亟需融入智能化和智

能原生的新动能。智能化是应用AI技术对存量系统的智能改造, 这一概念相对容易理解。与之不同, “智能原生”的概念虽然已经被普遍接受^[3], 但其关注的是AI带来的智能增量, 且理念更为超前, 相关研究还处于初步阶段。对智能原生的探讨可以从技术生态、科技伦理、数字文明等不同维度展开, 也可以选取更具体的切入点。本文选择从技术、业务和人3个方面, 探讨并展望对智能原生的一些理解。

1 AI原生技术

AI原生技术指那些本身为AI而设计, 基于AI第一性原理而设计、开发并运行的技术体系。本文将聚焦AI基础设施(AI Infra)层技术, 仅探讨原生多模态、原生数据和原生算力, 以及与智算中心配套的供配电和制冷技术。

1.1 原生多模态

目前主流的多模态模型是多个单模态模型的简单拼凑。

这类模型通过独立训练文本、图像、语音等子模型，再在后期实现子模型间的协同对接以完成多模态交互。这种拼接式的多模态架构，使得模态间信息传递依赖于手工设计的特征工程，容易导致语义关联丢失，也限制了模型的跨模态推理能力，使其难以应对复杂的应用场景。

原生多模态模型首先需要进行架构层面的创新，以深度融合各种模态。比如，不同模态需要共享参数空间，构建统一的表征空间，将不同模态数据映射至同一语义维度；引入跨模态的注意力机制与模态感知的权重分配机制，使模型能够动态聚焦关键信息；采用多模态联合优化训练策略，而非独立训练子模型，具体可以采用分阶段训练策略：先固定语言模型参数，再优化视觉编码，最后完成全模型的微调。

原生多模态除了需要实现以上架构层面的创新外，还需要在数据层面构建多模态数据库、在算力层面研发多模态芯片，并在跨模态信息的精准对齐与交互、多模态工具链研发、人机交互与用户体验等方面，开展更多的原生性创新实践。

1.2 原生数据

2025年，大模型训练“撞上”了数据墙已经成为业界共识^[4]，即模型训练所需的高质量数据即将耗尽或难以获取，导致模型性能提升陷入停滞甚至出现倒退。互联网上可用于高质量训练的数据已接近开采殆尽，并且现有数据大多存在多样性偏差、内容过时或信息虚构等问题。与此同时，企业私域数据流通依然受阻，行业领域数据的版权壁垒也居高不下。

无论是公网数据、企业私有数据，还是行业专业数据，采集、清洗和保存这些数据的初衷，可能是为了信息传播、社交互动、风险控制或溯源审计等，但几乎都不是为了训练大模型。只是近年来，随着深度学习和大模型技术的发展，这些数据经过结构化重构、隐私增强和人工标注等处理后，被用于大模型训练，实现了二次开发和新价值释放。训练AI时缺乏“专业教材”的现象，使得AI训练领域正面临着转变数据原始意图时的供需错位、训练资料缺陷和成本高昂等挑战。

更深一层看，大模型所谓的深度学习，本质上并不是从物理世界直接获得或通过互动习得的一手知识，而是来自人类中介的二手数据，这类似于从“纸面知识”而不是从实践中习得知识的模式。AI的这种间接学习模式，会将其智能水平局限于人类认知的边界内。结合前面的讨论可以发现，目前AI训练实际上是基于人类提供的二手数据开展二次开发与利用。

为此，从短期来看，除了继续优化传统的数据“转型”之路外，还需要主动设计并开发面向AI的数据战略，即把模型训练作为数据生产的第一目标，重构数据全生命周期，形成数据、模型和应用的闭环生态。比如，在数据采集阶段，根据模型任务（如对话、图像生成）场景化定制数据采集方案，同步采集文本、图像、传感器等多模态数据，以避免后期模态拼接引发的语义断层问题；在数据清洗阶段，利用模型自身能力自动筛选数据和反向校正偏差；在数据保存阶段，像代码库一样管理数据集的版本迭代，在进行数据存储时就完成隐私增强保护，以避免后期再开展脱敏处理等冗余操作。

哈佛大学的一项最新研究发现^[5]，在大模型训练数据中加入约10%的毒性数据，可以使毒性特征更集中且更易分离，更有利模型的长期稳健性。这为业界解决“高质量”数据问题提供了新思路：与其付出极高代价对数据进行彻底清洗，可能还不如引入生物学上的毒性免疫原理。AI大模型的输出是统计意义上的最优解，即使引入极少量的噪声数据（比如错误数据、毒性数据等）对输出结果的影响也很有限。

为此，从长远来看，模型训练应该实现“去人类中介”，通过构建多模态直接感知体系，更直接地触及世界运行的本质规律。比如，通过高精度脑机接口捕捉神经元集群的原始电信号，利用分布式触觉传感器网络，直接感知物体表面的分子排列与力学特性，结合环境交互反馈形成闭环学习，让AI在物理世界中自主验证认知假设等。

1.3 原生算力

在AI发展的早期阶段，相关运算主要依赖通用中央处理器（CPU）进行简单计算，AI与芯片彼此之间几乎没有相互影响。后来随着深度学习的兴起，AI模型的参数规模开始激增（2012年，AlexNet模型的参数规模已达6 000万），通用CPU算力不足的问题开始凸显。而原来主要用于图形处理的图形处理器（GPU）凭借并行计算（包括软件部分的计算架构）的优势，被引入AI大模型的训练。但在引入初期，GPU仅被当作加速工具，经常未针对AI特性做深度优化。2017年之后，随着Transformer架构的提出、大规模模型训练的普及以及缩放定律（Scaling Law）的发现，AI模型的复杂度和算力需求呈指数级攀升。各类芯片厂商针对AI特性开始深度优化硬件架构，让原本主要服务于图形处理的GPU，快速演变成了以AI为核心用途的算力模式。

AI与算力的关系，已从早期的通用支撑、中期的辅助性加速器，发展到现在的原生芯片创新与AI算法优化循环

的共生关系。AI原生芯片针对AI算法的特性做了大量的定制化开发和深度优化。比如，在架构方面，原生算力采用数据流架构和脉动阵列等，减少内存搬运，进而提升计算效率；在指令集方面，设计张量指令集以支持矩阵乘加，扩展混合精度与动态量化指令，以适配AI计算特性；在通信协议方面，应用远程直接内存访问（RDMA）等技术实现CPU旁路与直接内存访问等；在精度优化方面，采用FP8低比特训练技术，降低计算与显存开销。

1.4 与配电制冷的协同

AI数据中心（AIDC）相较于传统的互联网数据中心（IDC），除了需要配备完善的IT基础设施外，还需要特别关注电力和制冷系统。这是因为，AI算力需求激增会导致硬件功耗和散热大幅上升。传统IDC主要以通用CPU为主，单机柜功率为4~8kW，采用风冷方式即可满足散热需求。而AIDC依赖GPU和张量处理器（TPU）集群，单机柜功率通常可达10~100kW，部分高密度场景甚至更高。高密度算力下硬件易因过热降频或损坏，使得液冷成为了刚需。AI原生的液冷，已经形成了冷板式与浸没式两大技术路径。

2 智能原生的业务

各类技术的应用演进一般都会经历原生应用、技术化转型与技术性原生3个阶段。

过去20余年，互联网应用的转型和原生演进过程契合这一规律。在互联网信息展示方面，Yahoo公司的门户网站，本质是传统电话黄页的数字化翻版，用户还需主动逐级检索；Google公司的搜索引擎，则通过算法对网页相关性进行概率排序，以关键词关联链接实现高效触达，原生出了对用户意图的动态理解。在互联网通信方面，IP电话和基于IP的语音传输（VoIP）通过IP协议改造传统语音传输，本质仍是电话服务的数字化延伸；而社交网络基于互联网的即时性与社交属性，原生出了人际交互的新范式，彻底摆脱了传统通信以话音为中心的模式和单纯工具属性。在互联网媒体方面，网络电视（IPTV）以IP网络传输电视信号，但内容生产仍沿袭传统电视的遥控器操作、频道编排和节目播控模式。而如今互联网上的视频服务模式，原生出短剧、竖屏与弹幕互动等的典型特征，这在20多年前讨论三网融合时是难以想象的。

从互联网时代发展到人工智能时代，AI与各行业的融合也将会是一个逐步演进的过程，且大致包括“智能化转型”和“智能原生”两种形态。其中，智能化转型是AI技

术对传统业务的渐进式赋能，其核心目标是通过算法优化并提升现有业务的业务效率、降低成本，相关技术多作为工具嵌入业务流程。智能原生是将AI技术深度融入业务基因后的范式变革，其核心特征是以AI技术为基础（或以AI为核引擎），实现对现有业务的深度改造，甚至创造出全新业务形态。

当前，已经出现了一些AI原生的新应用。例如，AI编程颠覆了传统手动编码范式，借助自然语言交互实现高效优质的代码生成；硅基人突破了预设脚本的限制，在数字空间中完成高保真孪生建模、动态人格重塑与沉浸式化身创造；AI编剧打破传统内容生产的线性流程，支持主题驱动的自动化剧情生成与实时叙事调整；AI通信并非简单地把AI技术叠加到传统通信之上，而是通过脑机接口等实现思维信号的无损传输与意图直连；AI原生企业则将人工智能深度嵌入企业运营全链路，构建智能决策中枢、持续创新引擎与个性化价值创造网络三大核心能力。

3 AI原生代

人是决定技术演进与社会发展的核心因素。计算机和互联网等数字技术把人群分化成数字难民、数字移民和数字原生代等，而AI正在把人群的分化从数字化层面推升级至智能化层面，催生出对应的AI难民、AI移民和AI原生代等。

AI原生代天然会将把互联网和云计算等当作传统基础设施而非新型基础设施。他们将伴随着大模型、智能体（Agent）和具身智能等智能原生新业态共同成长，在社交模式、消费逻辑和生命伦理等方面的理解，即便与传统数字原生代相比，也将呈现出显著差异。当然，AI原生代还必须面对一个全新的挑战，即如何定义碳基生物人与硅基AI人之间的关系。

在社交模式方面，AI原生代将不再局限于线上人与人之间的社交关系，而会进一步扩展到与“AI人”（也称数字人、数智人、数字分身、数字化身、硅基人等）打交道、当朋友、谈生意甚至谈恋爱等。互联网实现了“信息在线”，转移了数字原生代的注意力，并且发展出了一套注意力生态；AI人将实现“情感在线”，会转移AI原生代的情感依托，并且发展出情绪价值的新业态。未来，硅基人很有可能会超越传统社交网络中的碳基人类，为AI原生代和AI移民提供情感陪伴、会话聊天、情感支持甚至参与决策。这将对现代社会的社会伦理秩序和婚姻家庭制度带来全新挑战。

在消费逻辑方面，随着全球物质生产的极大丰富，人类社会正在从以物质消费为核心，过渡到以精神消费为重点的新消费时代。AI原生代尤为关注情绪价值的消费，或者说

属于“情绪价值优先”的一代人。AI原生代将通过算法筛选和推荐，寻找能够满足自身高阶情感需求的AI产品和服务，例如为虚拟偶像消费、购买具有情感共鸣的数字藏品等，推动情绪经济向规模化、专业化方向发展。

从自我认知方面，基于AI的数字分身会使得原生代的身份认同变得更加复杂。传统数字原生代通常随身携带一部智能手机，将其作为连接世界的窗口，以满足信息获取、社交互动和娱乐消费等需求。AI原生代通常会配备一个AI助理（物理形态不限于手机，可以是各类可穿戴设备），为其提供主动服务和智能决策。早期的数字移民生活在一个线上线下相分离的世界，传统数字原生代生活在一个线上线下相结合的世界，而AI原生代将生活在一个线上线下深度融合、真实与虚假共存、碳基智能与硅基智能并列的新世界里。

4 结束语

本文从技术、业务和人3个角度对AI原生展开初步探讨，希望能够对AI的应用落地和模式创新起到一定的参考作用。另外，与互联网应用和云原生概念类似，对智能原生的理解和实践也将是一个不断发展演进和丰富的过程。

参考文献

- [1] IDC. 全球数字化转型支出指南 [EB/OL]. [2025-11-05]. https://my.idc.com/getdoc.jsp?containerId=IDC_P32575
- [2] 何宝宏. 数字原生 [M]. 北京: 中译出版社, 2023
- [3] 中国政府网. 国务院关于深入实施“人工智能+”行动的意见 [EB/OL]. (2025-08-25) [2025-11-05]. https://www.gov.cn/zhengce/zhengceku/202508/content_7037862.htm
- [4] Economist. AI firms will soon exhaust most of the internet's data [EB/OL]. (2024-07-27) [2025-11-05]. <https://www.economist.com>
- [5] LI K, CHEN Y, VIEGAS F, et al. When bad data leads to good models [EB/OL]. (2025-05-07) [2025-11-05]. <https://www.arxiv.org/abs/2505.04741>

作者简介



何宝宏，中国信息通信研究院云计算与大数据研究所所长；长期从事互联网、云计算、大数据和人工智能等领域的技术研发、标准制定与产业研究工作。



智算光互联技术进展及趋势

Advances and Evolutionary Trends in Optical Interconnection Technology for Intelligent Computing

张平化/ZHANG Pinghua¹, 王会涛/WANG Huitao¹,
王勇/WANG Yong²

(1. 中兴光电子技术有限公司, 中国南京 210012;
2. 中兴通讯股份有限公司, 中国 深圳 518057)
(1. ZTE Photonics Technology Co., Ltd., Nanjing 210012, China;
2. ZTE Corporation, Shenzhen 518057, China)

DOI: 10.12142/ZTETJ.202506008

网络出版地址: <https://link.cnki.net/urlid/34.1228.TN.20251219.0951.002>

网络出版日期: 2025-12-19

收稿日期: 2025-10-16

摘要: AI大模型的快速迭代,直接推动集群算力需求进入高速增长通道,而互联技术作为破解集群算力瓶颈的核心抓手,其战略价值不断提升。智算集群更是将互联需求从“单纯扩规模”推向“规模、性能、灵活性三位一体升级”。从技术底层逻辑出发,介绍了智算光互联关键技术进展,认为光电融合的核心是光互联技术的应用边界持续向芯片级、网络级延伸。而光输入输出(OIO)/共封装光学(CPO)互联技术与光交换(OCS)技术,正是精准适配这一发展趋势的关键技术方向——既能满足智算集群对高带宽、低能耗、高可靠的核心需求,又可兼顾多租户场景下资源灵活分割与调配的互联诉求。

关键词: 光电融合; 共封装光学; 光输入输出; 光交换

Abstract: The rapid iteration of AI large models has directly propelled the demand for cluster computing power along a high-growth trajectory. As a core lever for breaking through the bottlenecks in cluster computing power, interconnection technology is witnessing a steady rise in its strategic value. Particularly for intelligent computing clusters, interconnection requirements have evolved beyond "simple scale expansion" to "a trinity upgrade encompassing scale, performance, and flexibility". From the perspective of underlying technical logic, the key technological advancements in intelligent computing optical interconnection are discussed. It is believed that the core of optoelectronic fusion lies in the continuous extension of optical interconnection technology's application boundaries—expanding both to the chip level and network level. Notably, optical input/output (OIO)/co-packaged optics (CPO) interconnection technologies and optical circuit switching (OCS) are precisely the key research directions that align with this development trend. They meet the core requirements of intelligent computing clusters for high bandwidth, low energy consumption, and high reliability, additionally accommodating the interconnection needs for flexible resource partitioning and allocation in multi-tenant scenarios.

Keywords: optoelectronic fusion; CPO; OIO; OCS

引用格式: 张平化, 王会涛, 王勇. 智算光互联技术进展及趋势 [J]. 中兴通讯技术, 2025, 31(6): 53–60. DOI: 10.12142/ZTETJ.202506008

Citation: ZHANG P H, WANG H T, WANG Y. Advances and evolutionary trends in optical interconnection technology for intelligent computing [J]. ZTE technology journal, 2025, 31(6): 53–60. DOI: 10.12142/ZTETJ.202506008

1 AI时代的光互联需求

AI大模型参数规模快速增长,推动算力需求进入指数级增长通道(每两年增幅达100倍),但硬件性能提升滞后(单颗图形处理器算力每两年仅增3.3倍,内存及电互联带宽增幅更低),导致单一计算单元无法支撑大模型高效训练,成为产业发展梗阻^[1]。为填补算力缺口,业界采用集群化方案:先将多颗图形处理器(GPU)整合为超节点,再通过超节点组网构建AI集群。这一方案对互联技术提出“低时延、高带宽、低损耗”的要求,而传统电互联因物理局限难以适配。具体表现为:在传输性能方面,印制电路板(PCB)与铜缆插入损耗高、信号衰减快,限制跨板传输和

架构扩展;在带宽与集成效率方面,存在带宽天花板,边缘带宽密度低,制约集成效率;在部署与能耗方面,铜缆用量大,使设备重量、部署复杂度及能耗显著增加,与轻量化、低功耗目标相悖。

行业分析指出:未来10年,不是算力增长跟不上需求,而是若没有光技术的突破,算力需求将永远被“传输瓶颈”所束缚。若缺乏光技术的突破性应用,这种“传输瓶颈”会像无形的锁链一样,将海量算力困在“数据无法高效流动”的牢笼中——一边是集群内闲置的算力资源,一边是大模型训练的算力饥渴,最终形成“算力过剩”与“需求饥渴”并存的荒诞局面。而光技术,正是打开这一牢笼的唯一钥

匙。如图1所示,光输入输出(OIO)、共封装光学(CPO)、光交换(OCS)作为光互联技术落地的三大核心载体,正分别在“芯片级”“封装级”“集群级”3个关键维度精准破局:OIO依托封装内微光子器件,实现芯片间微秒级的极致低时延互联;CPO通过将光引擎与芯片封装集成,解决封装内信号损耗与能耗过高的问题;OCS以光信号交换替代传统电子交换,化解万卡级集群的带宽拥堵难题。三者共同构建“光-电-算”深度协同的算力流动新体系,为AI算力的高效释放提供全链路技术支撑。

2 芯片级光互联OIO

如图1所示,外部光信号经光纤阵列进入OIO芯粒完成“光→电”转换,电信号传输至中间算力芯片进行计算后,再由OIO芯粒转为光信号。随后,光信号经另一侧光纤阵列输出。这一过程实现了“光互联+电计算”的深度融合,为超大规模AI集群、全光算力网络等场景,提供高带宽、低时延的互联与计算支撑。

2.1 OIO实现方案

AI算力爆发推动芯片级光互联OIO技术向“高集成、低功耗、广适配”方向升级。在这种背景下,不同厂商依托自身技术积累、核心场景定位和供应链优势,形成了差异化显著的核心技术方案,如表1所示。

1) SiGe EAM方案

如图2所示,SiGe EAM方案是以SiGe电吸收调制器(EAM)为核心的光互联解决方案,依托TSMC 5 nm制程实现电子、光子系统深度协同。

SiGe EAM作为光子芯片(PIC)的关键发射端器件,其结构紧凑(调制区长度<50 μm)且热稳定性优异(-40~85 $^{\circ}\text{C}$ 下调制效率波动<5%)。该方案需与TSMC 5 nm电子芯片(EIC)深度配合:EAM将EIC放大、均衡后的电信号,编码到外部激光源(1310 nm/1550 nm)的激光载波上,信号经SiPh波导低损耗传输后,通过光栅耦合器与光纤阵列单元(FAU)批量耦合至外部光纤;在接收端,PIC内部的光电二极管将光信号转为电流,信号经EIC放大、解串后回传至电子系统。

该方案通过PIC中介层/光学多芯片互联桥(OMIB)与铜柱凸点实现EIC与PIC高密度异构集成,兼具低功耗(动态功耗为6.08 fJ/bit)、互补金属氧化物半导体(CMOS)工艺兼容优势,为数据中心AI训练、超算集群等场景提供Tbit/s级带宽、小于1 ns延迟的高效光互联支撑。

2) μLED 方案

如图3所示, μLED 方案(以Avicena Light BundleTM为代表)是基于微型发光二极管阵列的高密度光电异构集成互联技术,通过 μLED 阵列、多芯光纤束、微透镜阵列、PD阵

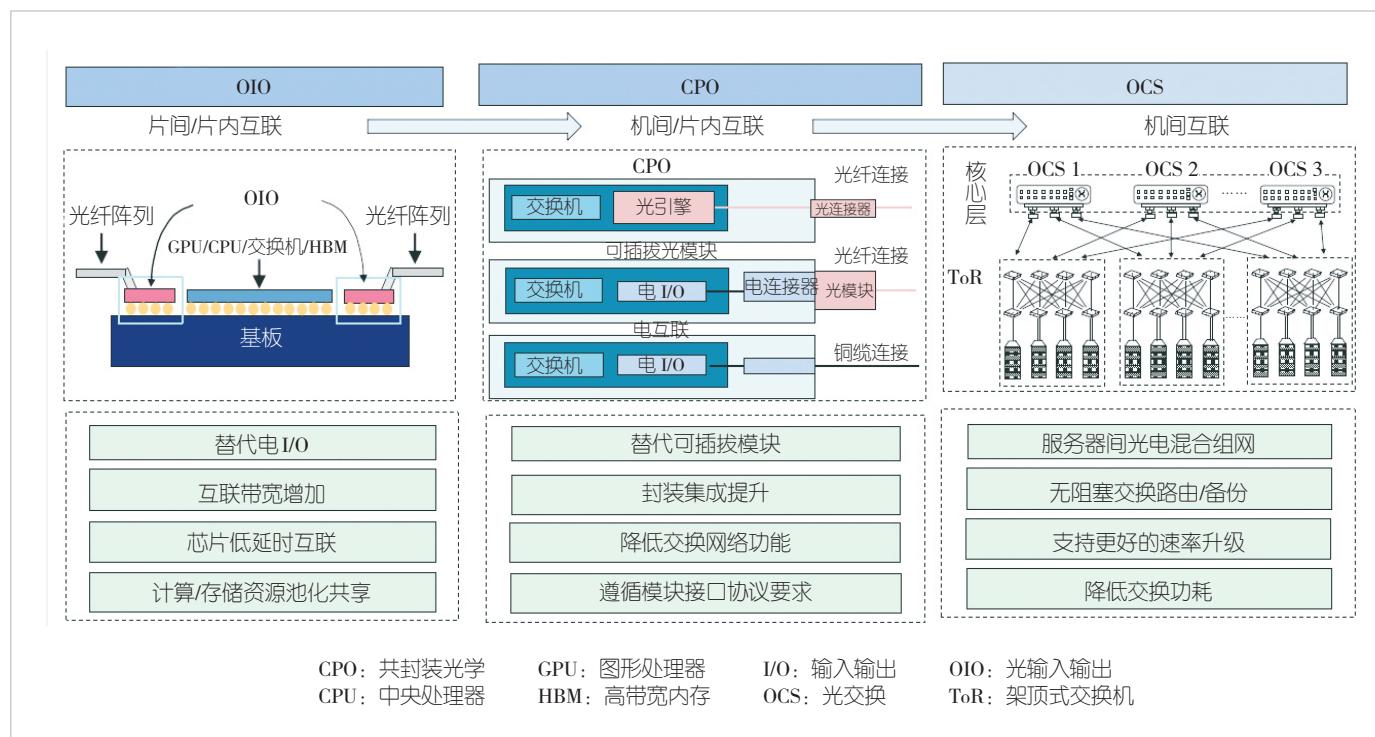
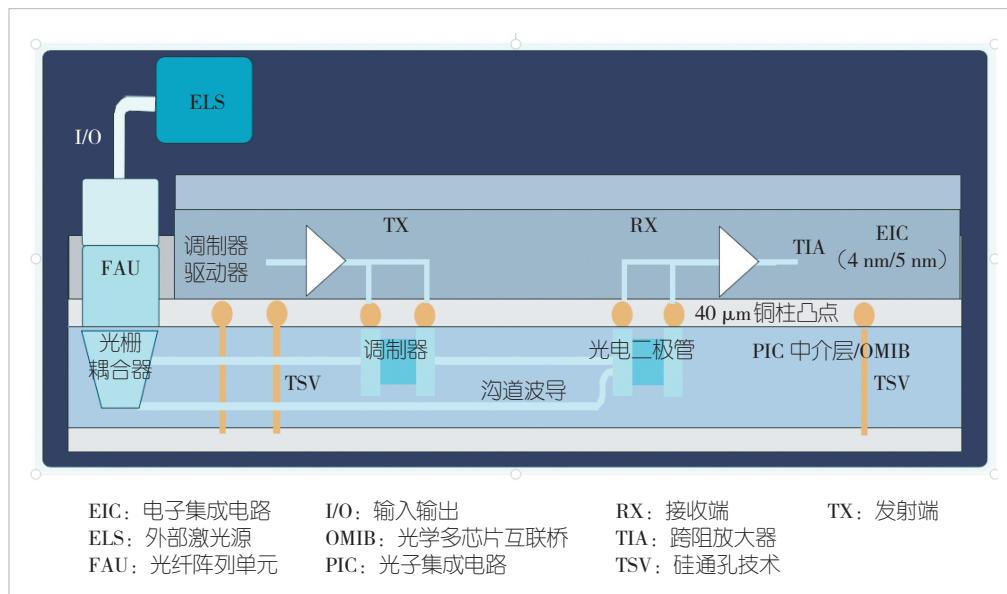
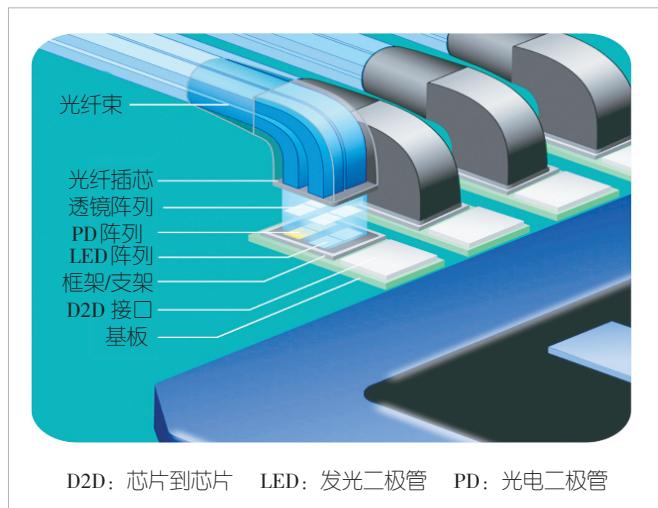


图1 光互联三大关键技术

表1 主流厂商OIO技术方案对比

厂商	核心方案	核心优势	应用挑战
Celestial AI	SiGe EAM 方案	1. CMOS 工艺兼容性: SiGe EAM 的核心工艺(如外延生长、光刻、刻蚀)与硅基 CMOS 完全兼容 2. 宽温稳定性: 硅基材料的温度免疫革命 3. 功耗优势: 基于电场调控吸收系数的电吸收效应, 静态功耗趋近于零, 动态功耗仅源于信号驱动	1. SiGe 外延工艺复杂性高: 性能高度依赖外延层质量, 生长时需精确控制锗含量与应变平衡, 工艺控制难度大 2. 非线性响应需专用补偿: 调制曲线存在非线性特性, 需通过专用技术进行补偿以保证信号质量
Avicena	μ LED 方案	1. 超低能效比: <1 pJ/bit, 大幅降低互能耗 2. 成本优势: 复用显示产业链, 降低商业化门槛	1. 单波速率受限: 单波速率局限于 10 Gbit/s 2. 材料工艺瓶颈: GaN 外延缺陷与波长均匀性难题
Ayar Labs/ Intel	MRM 方案	1. 超紧凑尺寸: 微米级几何突破, 芯片级光互联 2. 波分复用: 从单通道到多维度集成 3. 超低功耗: 电容负载优化+硅基工艺兼容, 实现皮焦级能效	1. 温度敏感性: 微环的调制性能高度依赖输入波长与谐振波长的匹配度, 温度导致的波长漂移会引发连锁反应 2. 非线性失真: 硅基载流子效应导致的信号质量劣化

CMOS: 互补金属氧化物半导体 EAM: 电吸收调制器 MRM: 微环调制器 μ LED: 微发光二极管

图2 SiGe EAM示意图^[2]图3 μ LED方案示意图^[3]

列以及芯片级电互联组件的协同, 构建“电-光-电”高效转换与长距传输链路。

3) MRM 方案

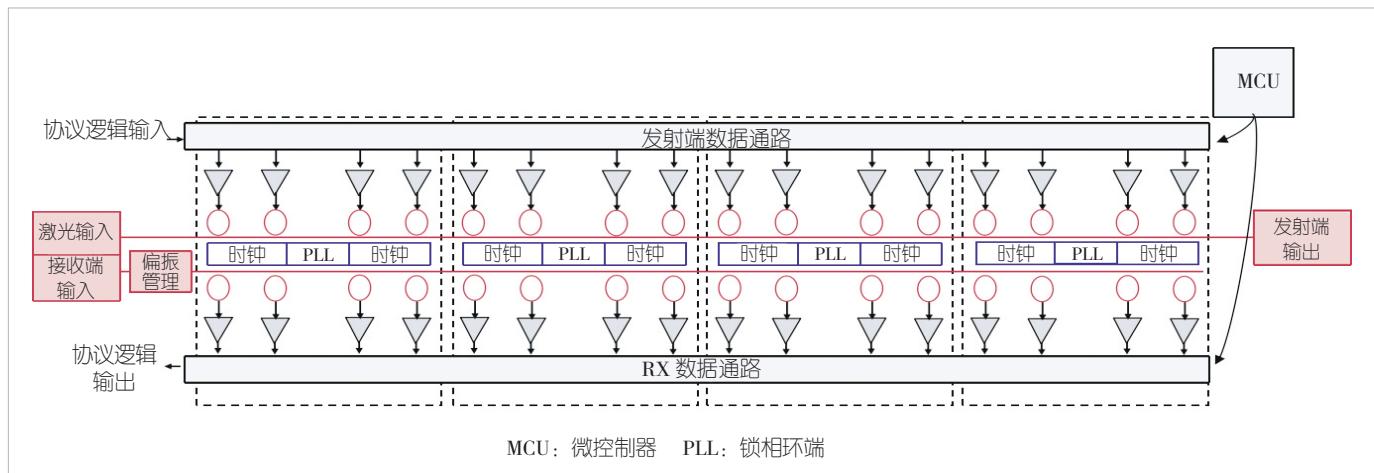
如图4所示, MRM 方案是高速光互联核心架构, 通过接收端 (TX) 数据通路、发射端 (RX) 数据通路、锁相环 (PLL)、微控制器 (MCU) 深度协同, 构建集高速信号调制、偏振补偿与智能管控于一体的高效传输机制。

在 MRM 方案中, 发射链路完成“电→光”高精度转换: 接收电信号后, 经内置时钟网络时序校准, 结合激光输入调制, 从输出端输出高稳定性光信号。接收链路聚焦“光→电”高保真还原: 通过偏振管理模块补偿单模光纤传输中的偏振态漂移, 经时钟网络重定时后, 将光信号还原为电信号回传协议逻辑。

架构上, MRM 方案采用“4组通道共享 PLL 与时钟资源”策略, 降低功耗并减少芯片面积。片上 MCU 实现全流程智能控制, 支持每通道 32 Gbit/s 非归零码 (NRZ) 高带宽传输, 凭借低损耗、高稳定性成为 AI 芯片互联与数据中心光通信的核心支撑技术。

2.2 OIO发展趋势

OIO 技术演进的核心是实现光电融合架构从“电互联辅助层”向“算力基础设施核心架构”的范式跃迁。这一跃迁

图4 MRM方案示意图^[4]

需要以OIO技术突破系列核心瓶颈为前提——这些瓶颈既涉及短期AI集群落地的技术适配与稳定性问题，又涉及中期生态标准化、量产规模化的诉求，同时也决定了算力基础设施全光化重构的长期发展。因此，瓶颈突破路径需与OIO三阶段演进深度适配，具体按技术阶段拆解如下：

- 1) 短期（AI集群落地验证阶段）：推动“外置光源+硅光引擎”混合架构在AI集群场景中实现稳定可靠运行，实现技术适配与工程化验证的闭环。
- 2) 中期（规模化普及阶段）：推动OIO技术规模化落地，核心是让技术具备商业化普及的经济性——一方面通过统一行业标准降低下游客户的适配成本，另一方面借助量产规模效应降低硬件研发与制造成本。当前需重点攻克与这两大目标对应的核心难题。
- 3) 长期（底层重构阶段）：对现有算力互联架构进行底层重构，最终目标是落地“晶圆级光芯片”与“全光化算力架构”，实现算力传输与管控的全链路光化升级。当前需优先攻克两类具有颠覆性的技术难题，为目标落地扫清关键障碍。

3 封装级光互联CPO

CPO技术的核心是打破传统芯片-光模块-主板分离式架构，将光引擎（集成调制器、光探测器等）与芯片（高端交换机芯片、AI加速器芯片等）共同集成于同一封装基板，形成“芯片-光引擎”一体化封装单元^[5]。其技术定位介于芯片级OIO（封装内光接口）与板级光互联（芯片通过线缆连接外部光模块）之间，将光信号传输距离从传统方案的厘米级（板级布线）缩短至毫米级（封装内互联），在物理层面解决光互联的核心痛点。

3.1 CPO实现方案

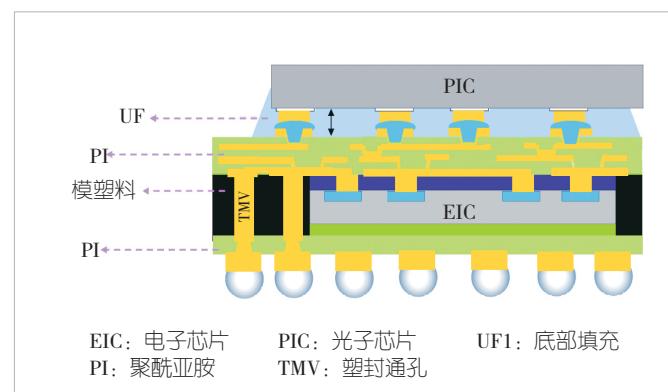
如表2所示，CPO作为突破数据中心高速互联瓶颈的核心技术，已形成两大成熟且具有代表性的主流实现方案，分别以博通（Broadcom）和英伟达（NVIDIA）为主导。二者在封装架构、光电子集成方式及核心光调制器件选择上存在显著差异。

1) Broadcom方案

如图5所示，Broadcom在FOWLP-2.5D架构下推出的“PIC-on-EIC”方案。该方案以“2.5D封装+光引擎上置”为核心架构逻辑，深度发挥扇出型晶圆级封装（FOWLP）技术的集成优势：先通过FOWLP工艺构建统一的高性能封装基板（材质可选有机基板或陶瓷基板，兼顾成本与散热需

表2 主流厂商CPO技术方案对比

对比维度	Broadcom	NVIDIA
封装架构	2.5D封装	3D封装
集成方式	光子芯片置于电子芯片上方	电子芯片置于光子芯片上方
核心调制器	马赫-曾德尔调制器	微环调制器

图5 Broadcom的CPO方案示意图^[6]

求), 该基板不仅承担机械支撑作用, 还能通过内置的高密度布线实现多芯片间的信号互联。随后采用“上层光、下层电”的堆叠布局, 将集成了马赫-曾德尔调制器 (MZM) 的 PIC 直接倒装或贴合于 EIC 之上, 形成垂直紧凑的异构集成结构, 大幅缩短 MZM 与 EIC 驱动电路间的物理距离。该方案通过统一封装体集成, 有效解决了传统光电器件多组件分离带来的兼容性差、装配复杂等问题。该方案的优点具体表现为: 探测器与 EIC 近距离集成, 减少光信号长距离传输的衰减; 搭载 MZM 的 PIC 与 EIC 垂直堆叠, 使电驱动信号可直接高效作用于调制器, 避免外部布线的干扰与延迟。最终在实现封装体积“轻量化”的同时, 提升了整套光互联系统的传输效率、稳定性与可扩展性。

2) NVIDIA 方案

如图 6 所示, NVIDIA 方案以微环调制器 (MRM) 的“电-光调制”功能为核心, 依托 TSMC-SoIC™ 键合技术实现异构芯片的高密度集成, 通过组件协同解决光信号传输损耗与稳定性差的问题:

图 6 (a) 展示了 TSMC-SoIC™ 面对面 (F2F) 键合方式: 处理电信号的 EIC 与搭载 MRM 的 PIC 采用“有源面相对”直接键合, 键合间距小于 $10 \mu\text{m}$, 大幅缩短信号传输路径, 降低衰减, 提升了电-光交互通同步性。相比于传统键合方式, 这种方式的连接精度提升超 50%, 集成密度提高 16 倍以上, 为扩展带宽奠定基础。

图 6 (b) 呈现了紧凑通用光子引擎架构 (COUPE) 的“组件协同生态”: 支撑硅作为载体, 兼具散热功能, 避免局部高温影响 MRM 精度; TSMC-SoIC™ 铜-铜混合键合实现 EIC 与 PIC 稳固连接, 构建低阻抗信号通道, 将驱动信号延

迟控制在纳秒级, 使互联损耗降低 85% 以上; 穿芯片过孔 (TDV) 构建垂直信号传输通道; 嵌入式微透镜、光栅耦合器 (GC)、金属反射器组成“光学优化链路”, 通过聚焦光信号、调控光路、适配通信波段等方法, 将光信号高效耦合至外部光纤, 使耦合效率提升 20% 以上, 增强了带宽稳定性。

3.2 CPO 发展趋势

当 NVIDIA 宣布 2025 年旗下 Quantum-X 交换机将全面导入 CPO 技术时, 一场重构光通信产业链底层逻辑的变革已然拉开序幕。这一技术跃迁不仅标志着光互联从“分立组装”向“系统级集成”的范式转移, 更引发了行业对核心竞争力重构的深度思考——CPO 如何在保持高密度、低功耗优势的同时, 缩短与可插拔光模块在成本控制、可靠性保障和维护便捷性上的差距。未来 3 年, 随着标准化落地与成本曲线下行, CPO 有望在 AI 超算、智算中心等核心场景中占据 50% 以上市场份额, 将真正实现与可插拔光模块的“同台竞技”。

4 集群级光互联 OCS

OCS 的核心价值在于从底层技术逻辑上打破传统电分组交换机的局限, 为高带宽、低延迟的网络传输筑牢基础。传统电分组交换机在处理光信号时, 必须依赖“光电-电光 (O-E-O)”转换流程——先将接收的光信号转为电信号 (O→E) 完成分组处理, 再转回光信号 (E→O) 进行传输。这一过程不仅会造成信号损耗, 还会叠加单次 10~50 ns 的延迟, 更受限于电信号的处理速率。而 OCS 可直接在光域完成光路调度, 彻底跳过 O-E-O 转换环节, 在物理层精简信号的处理步骤, 从根源上弥补了传统方案的不足。在此基础上, OCS 进一步形成成本、功耗、速率兼容性三重核心优势。

4.1 OCS 实现方案

OCS 作为全光通信网络中的核心交换设备, 基于光交叉交换核心原理, 能够在 P 个输入光端口与 M 个输出光端口组成的端口矩阵中, 实现光信号的灵活动态切换与路由调度。其核心价值在于借助光域直接交互, 使服务

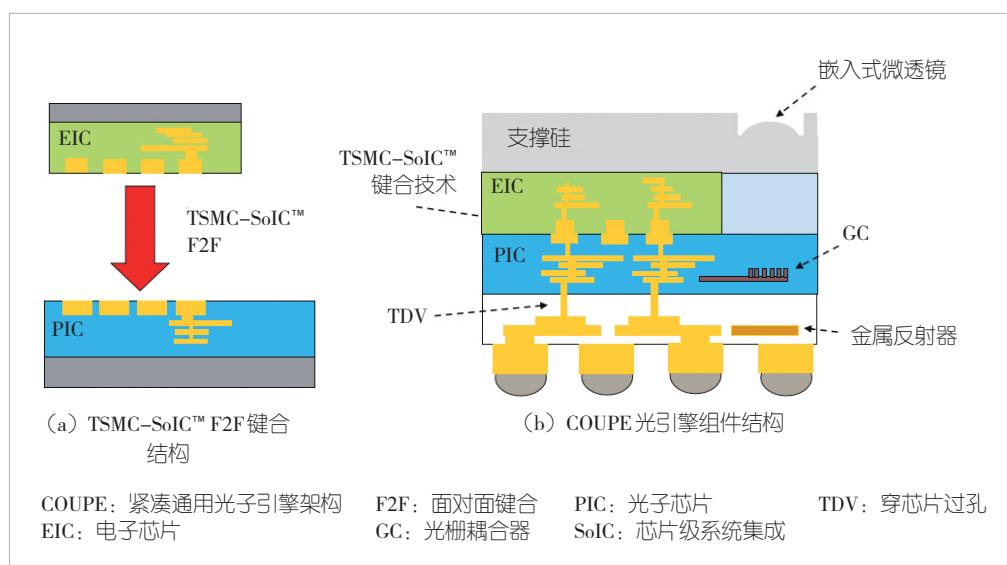


图 6 NVIDIA 的 CPO 方案示意图^[7]

器端口之间建立直达的光互联链路，完成全光路的实时重构与资源调度。如表3所示，当前支撑OCS落地的主流技术路线各有侧重，具体包括：以微镜阵列偏转实现光信号路由的微机电系统（MEMS）方案、依靠压电陶瓷驱动完成光路对准的压电陶瓷驱动方案、利用液晶电光效应调控光信号传输方向的液晶调控方案，以及结合硅光子技术实现高密度集成的硅光集成方案。这4类方案在性能、成本及适用场景上形成互补，共同推动OCS在数据中心、AI集群等场景的规模化应用。

1) MEMS方案

如图7所示，MEMS方案的原理为：输入/输出光信号经二维光纤准直器阵列（由 $N \times N$ 光纤阵列与二维透镜阵列组成）进入光学核心，核心内部的两组MEMS反射镜阵列通过驱动倾斜，引导光信号至对应端口。光路具备宽带、双向可逆特性，支持OCS实现数据速率无关的双向通信。此外，叠加的850 nm监控通道通过MEMS阵列发光、相机模块接收反馈，由控制硬件和固件优化反射镜驱动，以最小化信号损耗（每组MEMS阵列配一对发射与相机模块）。MEMS微镜阵列采用深反应离子刻蚀（DRIE）等方式实现精密工艺制造，要求角度校准精度 $\pm 0.001^\circ$ 、镜面粗糙度 $<10 \text{ nm}$ 。然而，这种方案也存在明显不足：在材料加工方面，硅材料应力梯度导致微镜翘曲（100 μm 边长微镜最大形变5 μm ），使晶圆级量产角度一致性偏差达 $\pm 0.01^\circ$ （超设计要求10倍），引发耦合效率下降、串扰增大的问题；在可靠性方面，微镜需耐受 10^{15} 次的切换（等效20年运行），但机械结构易受振动、 $-40 \sim 85^\circ\text{C}$ 宽温域影响，出现疲劳形变，导致角度漂移、响应延迟，削弱长期稳定性。

2) 压电陶瓷方案

如图8所示，基于直接光束偏转（DLBS）技术的压电陶瓷光开关方案，以“固态驱动、无复杂机械结构”为核心优势，在中低速光交换场景中占据重要地位。该方案采用“准直器-驱动器一体化”设计：高精度光纤准直器与压电陶瓷驱动器刚性固定，组成高密度二维阵列。两个阵列对向

平行布设，构成光开关矩阵核心。工作时依托压电陶瓷逆压电效应，通过电压控制准直器实现微米级位移与 $\pm 0.001^\circ$ 精度角度微调，使输入/输出端口精准对准，光信号直线传输完成交换。这种方案在性能上具有突出优势：光信号无需中间光学元件介导，插入损耗 $\leq 0.3 \text{ dB}$ （远低于MEMS光开关1.5 dB典型值），适配高速传输；固态结构无机械磨损，满

表3 OCS技术方案差异化对比表

对比维度	MEMS方案	压电陶瓷方案	液晶方案	硅光方案
切换速度	毫秒级	亚毫秒级	毫秒级	亚微秒级
使用技术	MEMS技术	DLBS技术	DLC技术	MZI技术
代表厂家	Google、Lumentum、华为	Polatis	Coherent	iPronics
	DLBS:直接光束偏转 MEMS:微机电系统 DLC:数字液晶 MZI:马赫-曾德尔干涉仪	MEMS:微机电系统 OCS:光交换		

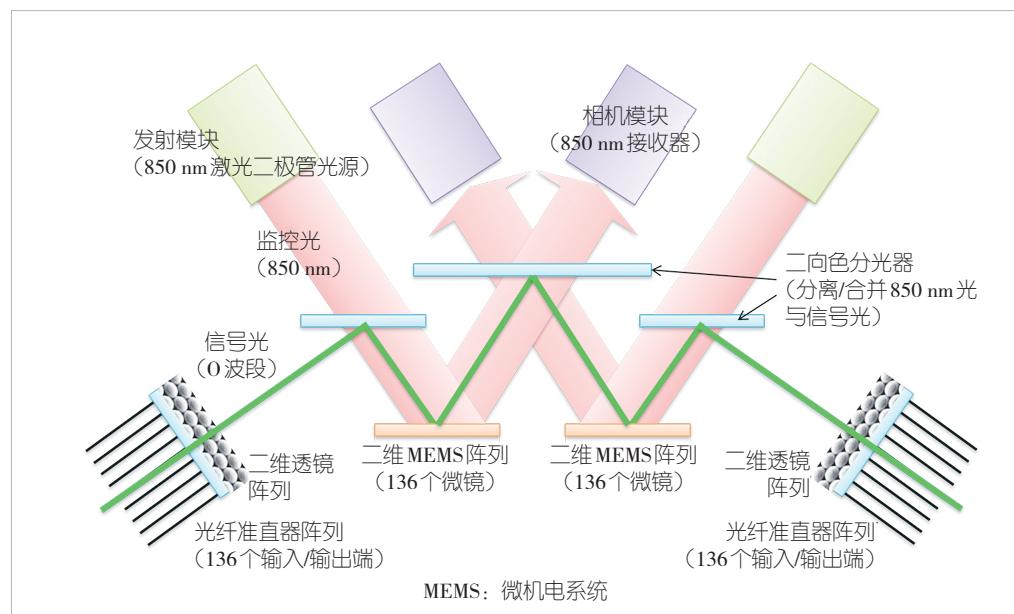


图7 MEMS方案示意图^[8]

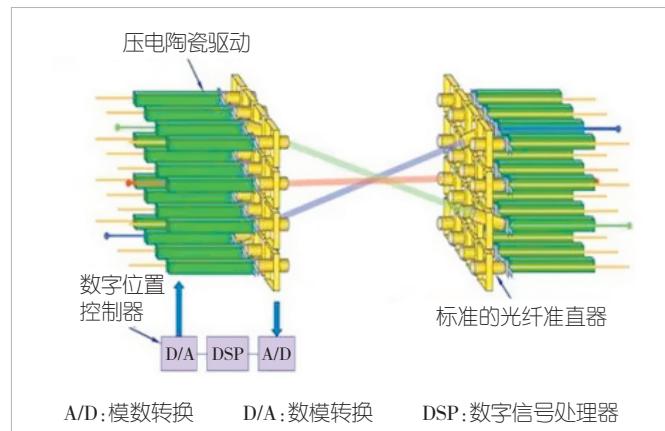


图8 压电陶瓷方案示意图^[8]

足高可靠性需求。但规模化应用面临瓶颈：矩阵向 256×256 及以上规模升级时，远距离交叉连接要求准直器具备更大转角与位移，对驱动器的驱动力矩和定位精度提出更高要求；同时阵列体积翻倍，部署成本增加，且间距拉大导致插损升至 1.2 dB 以上，影响了传输效率。

3) 液晶方案

如图9所示，液晶方案通过“输入光纤准直器阵列→偏振分光→X/Y方向晶体光楔与液晶二维偏转调控→输出偏振分光→输出光纤准直器阵列”闭环流程实现光信号动态互连：输入侧梯度折射率（GRIN）透镜型准直器将发散光转化为高平行度光束，经偏振分束器（PBS）分解为正交偏振分量；核心环节通过 0~5 V 电压调控液晶分子旋转，搭配双折射晶体光楔实现 $\pm 15^\circ$ 范围、0.01°精度的光束偏转；输出侧 PBS 整合偏振态后，由准直器耦合回光纤。该方案无机械部件，理论上插损 2~3 dB，寿命 $>10^8$ 次循环，且抗振动干扰。但产业化与大规模应用面临多重瓶颈：端口拓展受限，256 端口需 8 层级联（插损随层级递增），512 端口插损可能突破 5 dB，远超光网络“ ≤ 3 dB”的要求；响应速度 5~

20 ms，无法匹配 1~10 μ s 级实时调度需求；液晶折射率对温度敏感，需高精度温控模块（功耗增加 10%~15%）且易受电磁干扰，晶体缺陷还会加剧损耗差异；256 端口需大量偏振补偿元件，装配复杂度与多级级联损耗叠加，总损耗升至 5~7 dB，难以满足长距传输需求。

4) 硅光方案

如图10所示，硅光方案的核心基础单元为热调马赫-曾德尔干涉仪（热调 MZI），该单元依托硅材料的热光（TO）效应实现光信号的路径选择。MZI 单元内置两条平行的硅波导相位臂。系统向其中一条（或两条）相位臂的加热电极施加电流后，利用硅材料“温度变化→折射率改变→光程差变化”的特性，调控两相位臂中光信号的相位差，最终控制光信号从输出端口 O1 或 O2 导出。以该热调 MZI 为基础单元，图中 64 个同类器件通过交叉阵列式连接，共同构建 8×8 规模的硅光开关，可实现 8 路输入光信号与 8 路输出光信号的任意互联。在硅光方案中，随着端口规模的激增， 8×8 等小规模光交叉连接（OCS）的总损耗为 3~5 dB；而 64×64 及以上大规模硅光开关，因光信号需经过大量马赫-曾德

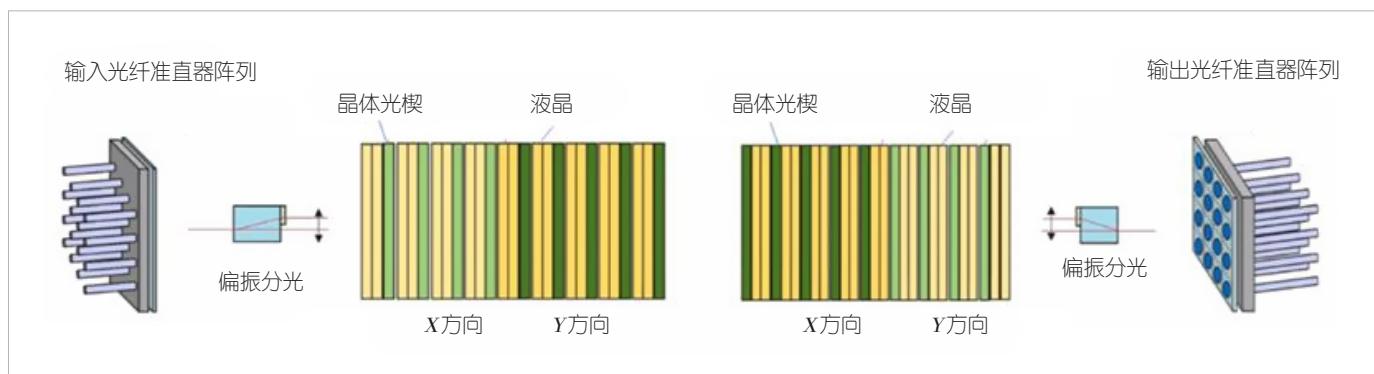
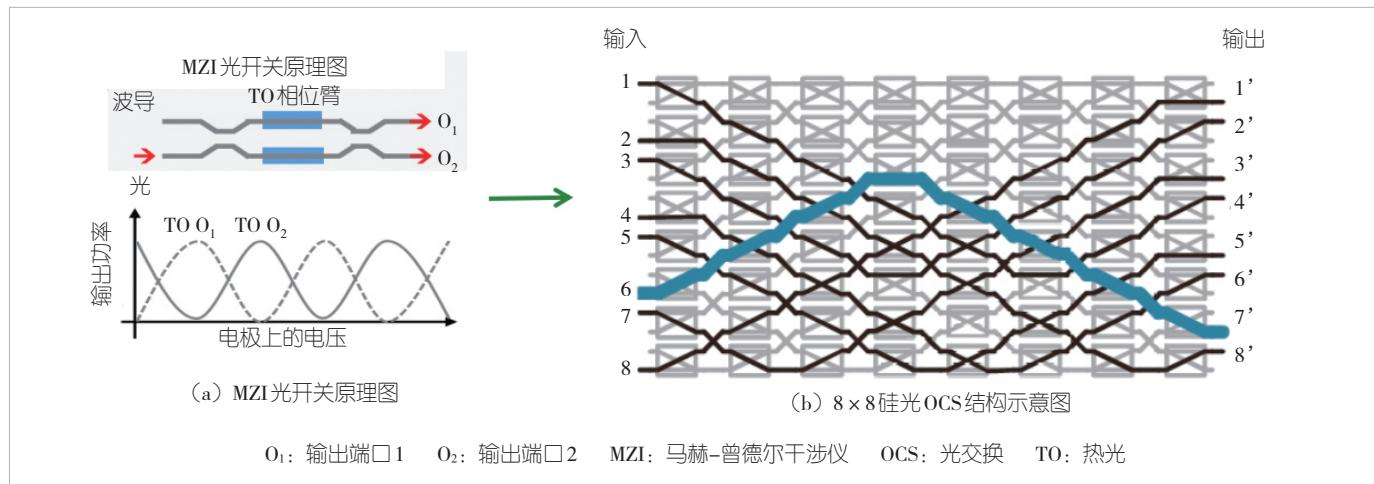


图9 液晶方案示意图^[8]



尔干涉仪（MZI）单元与交叉节点，传输损耗与元件损耗叠加后，总损耗显著超出光网络“≤3 dB”的核心指标要求。

4.2 OCS发展趋势

随着网络架构向超大规模全光化、AI驱动智能化、异构算力融合化加速演进，OCS作为未来网络核心支撑技术，面临三大核心要求：一是需支持单端口太比特每秒（Tbit/s）级带宽及千端口级无阻塞光交换，保障数据传输无瓶颈；二是具备AI驱动的自治式调度能力，与软件定义网络（SDN）协同实现毫秒级拓扑重构，适配动态业务需求；三是需突破能效与成本瓶颈，将单端口功耗降至亚毫瓦级、单个端口成本压缩至传统电交换设备的20%以内，为全光网络规模化部署奠定基础。

5 结束语

未来，随着光互联领域技术成熟度提升、行业互联互通标准统一及产业链协同深化，光互联技术将实现从“单点突破”到“体系化落地”的关键跨越。其中，CPO/OIO与OCS的协同演进，将彻底重塑数据中心网络架构逻辑，推动其突破电互联局限。二者将形成清晰的“分层协同”落地路径：CPO/OIO凭借“高密度集成”核心优势，聚焦数据中心“接入层”——通过芯片级/板级光互联封装，适配服务器、GPU集群的高带宽接入需求，成为接入层的核心解决方案；OCS则依托“全光调度”能力，立足“核心层+汇聚层”的架构中枢定位——通过连接不同机柜、机房的CPO/OIO设备，构建“无阻塞全光骨干网”。例如，在跨地域数据中心互联场景中，OCS结合C波段波分复用技术，可实现单条光链路数十太比特每秒（Tbit/s）级数据传输，且光路重构时间仅需5 ms，完美适配“东数西算”工程中“算力跨省调度”的动态需求。更重要的是，“CPO/OIO接入+OCS调度”的全光架构，彻底打破了传统“算力-数据-网络”的协同壁垒：过去依赖电交换完成的跨集群数据交互，如今可通过光域直接实现，不仅使算力资源利用率提升30%以上，还将AI大模型训练周期缩短近一半。在此变革下，数据中心将全面摆脱电互联的带宽、时延、功耗瓶颈，正式从“电联时代”迈入“全光互联时代”，为AI、云计算等数字经济核心产业的爆发式增长提供坚实支撑，并成为数字基础设施向“高效、绿色、智能”升级的核心引擎。

致谢

感谢中兴光电子技术有限公司沈百林、梁付运等专家对本研究的帮助！

参考文献

- [1] 韩炳涛, 刘涛. 大模型关键技术与应用 [J]. 中兴通讯技术, 2024, 30(2): 76–88. DOI: 10.12142/ZTETJ.202402012
- [2] Celestial AI. Photonic fabric module: the world's first system-on-chip (SoC) with in-die optical IO [R]. 2025
- [3] Avicena. Core technical advantage: best-in-class energy efficiency and bandwidth density [EB/OL]. [2025-11-06]. <https://avicena.tech/technology>
- [4] STOJANOVIC V. A UCLe optical I/O retimer chiplet for AI scale-up [C]// 2025 IEEE Hot Chips 37 Symposium (HCS). IEEE, 2025. DOI: 10.1109/HCS66204.2025.11154419
- [5] 沈百林, 王会涛. 相干光收发器件未来技术演进 [J]. 中兴通讯技术, 2023, 27(2): 82–86. DOI: 10.12142/ZTETJ.202305013
- [6] Broadcom. An AI compute ASIC with optical attach to enable next generation scale-up architectures [R]. 2024
- [7] CHEN M F, CHENG H T, TSOU C H, et al. Optical and electrical characterization of a compact universal photonic engine [C]// Proceedings of the IEEE 75th Electronic Components and Technology Conference (ECTC). IEEE, 2025. DOI: 10.1109/ECTC51687.2025.00015
- [8] 开放数据中心委员会. AI网络光交换机技术报告 [R]. 2025
- [9] SUZUKI K, KONOIKE R, CONG G W, et al. Strictly non-blocking 8×8 silicon photonics switch operating in the O-band [J]. Journal of lightwave technology, 2021, 39(4): 1096–1101. DOI: 10.1109/JLT.2020.3024016

作者简介



张平化，中兴光电子技术有限公司光模块规划工程师；主要从事光模块规划工作。



王会涛，中兴光电子技术有限公司规划总工；长期从事光通信系统、光电子器件产品和技术的研发及规划工作；曾获得中国通信学会科技进步奖一等奖、教育部技术发明奖二等奖；拥有专利10余项。



王勇，中兴通讯股份有限公司智算预研工程师；主要从事智算互联技术相关工作。



基于熵函数的语义信息边界研究

Research on Semantic Information Boundary Based on Entropy

唐雪/TANG Xue^{1,2}, 许进/XU Jin^{1,2}, 冯雨龙/FENG Yulong^{1,2}

(1. 移动网络与移动多媒体技术全国重点实验室, 中国深圳 518055;

2. 中兴通讯股份有限公司, 中国深圳 518057)

(1. China State Key Laboratory of Mobile Network and Mobile Multimedia Technology, Shenzhen 518055, China;

2. ZTE Corporation, Shenzhen 518057, China)

DOI: 10.12142/ZTETJ.202506009

网络出版地址: <https://link.cnki.net/urlid/34.1228.TN.20251219.1012.008>

网络出版日期: 2025-12-19

收稿日期: 2025-05-15

摘要: 人工智能的不断发展助力语义通信不断成熟, 语义通信越来越被证明是未来“智能体”之间通信的新范式, 然而语义度量理论发展缓慢且极不成熟。提出了语义度量的4条基本假设, 试图将经典信息理论与强语义信息理论等归入统一的语义度量理论中, 实现概率视角下语义度量方式与真性距离视角下语义度量方式的有机结合。在梳理语义度量理论发展历程的同时, 试图将其中的要素融汇贯通, 进一步给出相关概念, 并基于熵函数对语义信息的边界进行探索。

关键词: Shannon极限; 语义通信; 人工智能; 语义度量

Abstract: The ongoing advancement of artificial intelligence is facilitating the maturation of semantic communication, steadily establishing itself as a new communication paradigm for future “intelligent agents”. However, the development of theories for semantic measurement has been sluggish and remains underdeveloped. Four basic assumptions of semantic measurement are proposed, aiming to incorporate classical information theory and strong semantic information theory into a unified semantic measurement theory, and to realize the organic combination of semantic measurement methods from the probability perspective and those from the truth distance perspective. This paper attempts to explore the boundary of semantic information based on the entropy function, while combing through the development of semantic measurement theory and trying to integrate its elements.

Keywords: Shannon limit; semantic communication; artificial intelligence; semantic measurement

引用格式: 唐雪, 许进, 冯雨龙. 基于熵函数的语义信息边界研究 [J]. 中兴通讯技术, 2025, 31(6): 61–69. DOI: 10.12142/ZTETJ.202506009

Citation: TANG X, XU J, FENG Y L. Research on semantic information boundary based on entropy [J]. ZTE technology journal, 2025, 31(6): 61–69. DOI: 10.12142/ZTETJ.202506009

1948年, 香农发表“*A Mathematical Theory of Communication*”一文^[1], 高度概括了人类对于通信的认识, 指出数字化与编码是设计高效且可靠通信系统的必由之路。基于此, 通信行业经历了几十年的繁荣发展。直到今天, 经典通信的实现基本逼近香农理论极限, 通信技术的发展速度开始迟缓, 甚至不再聚焦于通信“本身”。通信领域的研究者提出的一部分新技术开始在复杂度、有效性和可靠性之间相互折中, 而另外一些新技术则依赖于材料、器件等方向的突破。就像19世纪的物理学一样, 可以说经典通信的大厦已经建成, 但其发展进程中仍存在两个关键的待解难题。

第一个难题源于信息论本身的假设条件。这一假设条件使其在实际通信系统搭建中存在很大的局限性, 不能面面俱

到地刻画人类的实际通信问题。原因有两点: 一方面人是智慧生物, 其交流的信息往往是非平稳的, 非各态历经的, 而概率论中没有有效的工具来刻画这类问题; 另一方面, 人类信息还具有模糊性, 比如文字的“韵味”或乐谱的“旋律”。其实, 在香农的论文出版之后, 当时的通信学家们就注意到了这个问题, 并被WEAVER总结为通信的三层问题: 语法问题、语义问题和语用问题^[2]。

第二个难题源于人工智能(AI)带来的不确定性。自人工神经网络(ANN)展现出强大的非线性拟合能力以来, 尤其是近期大模型对自然语言建模的能力, 这一特性更为凸显。ANN似乎为原本概率论中无法解决的问题带来了新的思路, 即采用ANN近似逼近^[3]。有人用流形理论解释ANN的这种非线性拟合能力。对于一个圆, 直接在一维坐标上表示是一群点的集合, 但是在二维坐标上可以用半径来表示, 将其投影回一维坐标, 则为一个点, 此时这个圆就是二维空

基金项目: 国家重点研发计划项目(2020YFB1807202)

间中的一维流形。通信数据无论是文本、图片、视频，大都是从人的角度，即三维空间建模的，但是 ANN 具有从更高维和更多维建模的能力，这就使得现实数据可以借助 ANN 在高维进一步表示、压缩和传输。而流形理论中的高维建模能力，其表现之一就是强大的非线性拟合能力。

随着 5G 技术的逐渐落地应用，国际电信联盟（ITU）发布了 6G 愿景示意图^[4]。未来 6G 强调数据处理的内容在不断增多，比如沉浸式通信、通感一体化等。相比于前几代通信，6G 愿景强调通信方面的内容在减少，唯一的新变化是 AI 与通信的融合。而语义通信起源于经典通信中的联合编码。联合编码则是最早试图采用 ANN 代替或增强经典通信系统的研究之一。因此，语义通信不仅与最早的通信研究有关，而且符合未来 6G 所重点关注的 AI 融合技术方向。但是，语义通信的理论基础还不够成熟，虽然它与经典通信同时提出，但是香农在对经典通信做了大量假设后，基于热力学中的熵函数得到了一套完备的描述体系。而语义度量理论面对的是非线性问题，即便想要转化为单纯的数学模型也极为困难。本文在梳理语义度量理论发展历程的基础上，尝试从熵函数的角度切入，得到关于语义度量的一些有用结果。

1 语义度量理论的早期发展

语义通信的研究主要沿 3 条技术路径并行推进。第一条路为基于香农经典信息理论（CIT）的发展。基于 CIT 的分离定理^[1]，当前的通信系统一般采用信源信道编码分离的方式搭建。但实际应用场景往往不满足分离定理成立的假设条件，因此人们试图测算联合编码方式能实现的增益，联合编码由此成为语义通信的开端。第二条路为语义度量理论的发展，包括 WEAVER 提出关于通信的三层问题^[2]、经典语义信息理论（CSIT）^[5]、强语义信息理论（TSSI）^[6-7]等，它们共同构成了语义通信的理论基础。第三条路为 AI 的发展历程，包括“图灵机”的提出^[8]、梯度回传算法的实现^[9]、深度学习^[10]以及现在的大模型^[11]等。语义通信主要是在联合编码系统的基础上，基于语义度量理论，利用 ANN 实现并发展的。如图 1 所示，语义通信架构在经典通信架构的基础上进行了扩展。本小节主要介绍语义度量理论的早期发展阶段和其中的主要思想。

1.1 针对通信的三层问题

CIT 的提出引发了学界的热烈讨论，并迅速形成两种对立观点。一派坚定地支持香农理论，认为其完备地描述了通信所面临的工程问题；另外一派则认为 CIT 存在极大局限，无法刻画人类实际通信背后的深刻意义。在这样的背景下，

当时的语言学大师 WEAVER 在 CIT 文章再版时，与香农讨论后为其补充了一段序言，将通信问题拓展为三层问题^[2]，一般将其翻译为：

- 语法问题：通信系统能多准确地传输符号？
- 语义问题：通信系统能多准确地传输符号含义？
- 语用问题：传输的符号含义能多准确地完成通信任务？

CIT 几乎完美地解决了语法问题。但受限于工程实现需求，CIT 将语义问题和语用问题排除在外，香农后来也曾在率失真理论文章中再次提到了人类对于符号的理解，即语义问题，称其无法通过简单的误码率来描述，而需涉及“可懂度”的概念。

语义问题的本质是建立语义空间到数据空间的映射，这一问题对应非线性编码过程，因此在统计理论中难以直接描述和求解。更进一步地，一般认为解决语义问题是解决语用问题的基础，只有理解了传输符号背后的含义，才能进一步将其映射为通信任务的完成程度。语义通信就是针对这两个问题背后复杂的统计理论而诞生的。可以认为，语义通信是语义问题和语用问题的综合解决方案。

1.2 经典语义信息理论

与 WEAVER 通信理论几乎同时诞生的是 BAR-HILLEL 与 CARNAP 提出的第一个对命题语义信息的描述^[5]，该理论一般被称为经典语义信息理论，简称为 CSIT。CSIT 沿袭了 CIT 的思路，但是将统计概率替换为语言的逻辑概率，并只针对一些逻辑命题。CSIT 认为一个命题包含的语义信息与其逻辑概率成反比，即一个命题出现的逻辑概率越大，其语义就越小。例如， $A = \text{“小白是条狗或猫”}$ ，其逻辑概率大于 $B = \text{“小白是条狗”}$ ，因而 A 的语义信息量就大于 B 。CSIT 定义的语义信息公式为：

$$\text{cont}(A) = \text{def} 1 - \text{Pr}(A) \quad (1),$$

$$\text{inf}(A) = \text{def} - \log_2 [\text{Pr}(A)] \quad (2),$$

其中， $\text{Pr}(A)$ 表示命题 A 出现的逻辑概率， $\text{cont}(A)$ 与 $\text{inf}(A)$ 分别表示描述命题 A 语义信息的两种形式。

但是，CSIT 的定义产生了一个悖论，人们称其为 BAR-HILLEL-CARNAP PARADOX (BCP)，即如果是矛盾命题，比如 $A \wedge \bar{A}$ ，由于其产生的逻辑概率为 0，那么按照 CSIT，其语义信息量为无穷大，这显然是违背常识的。BAR-HILLEL 与 CARNAP 也给出了相关解释，他们认为“矛盾本身就包含了太大的信息量”是其成为矛盾的原因，但这样的解释显然不能让人满意。之后也有很多学者沿着 BAR-

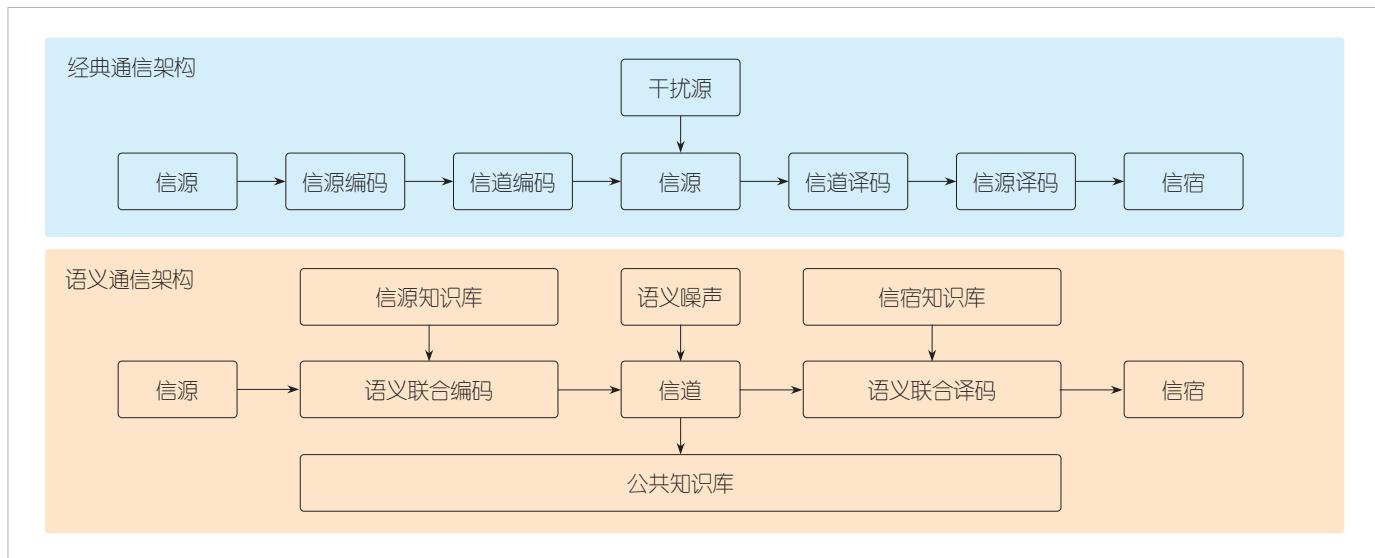


图1 经典通信系统架构与语义通信系统架构对比图示

HILLEL与CARNAP提出的研究路径描述信息量，但是都无法完美地解决这个悖论问题^[12]。

1.3 强语义信息理论

直到2001年，FLORIDI认识到BCP出现的原因在于没有规避语义真性的问题^[6]。在CIT中，由于人为规定了 $0 \log 0 = 0$ ，因此自然规避了不存在的信息。CSIT同样沿用了概率方法来描述语义信息，但是当命题语义不为真，即为矛盾命题时，就会出现语义信息趋于无穷大的情况。为了解决BCP问题，FLORIDI将语义真性封装在语义信息度量标准内，提出了另外一套语义度量标准^[7]，一般被称为强语义信息理论，简称为TSSI。TSSI中存在一个真命题，其他命题与真命题存在一定距离，该距离属于 $[-1,1]$ 区间。其中，-1代表全假，即该命题中没有一种情况符合真命题；1代表该命题符合所有情况的真命题，但是不止符合当前情况下这一种真命题；0则为当前条件下的真命题。语义信息公式定义为：

$$g(s) = 1 - f^2(s) \quad (3),$$

其中， $f(s)$ 表示命题 s 到真命题之间的距离， $g(s)$ 表示命题 s 包含的语义信息。

TSSI通过规定距离的方式定义语义信息，避免了BCP。但是由于FLORIDI本人是一位语言学家，其定义的语义信息计算公式过于唯象，公式之间缺乏推导和逻辑。这导致TSSI只适用于一些简单的逻辑命题。当命题之间需要进行逻辑运算，或者面对复合命题时，TSSI便难以有效处理。

1.4 语义真性与概率统计的结合

2011年，D' ALFONSO提出了一种将概率和距离融合起来的语义度量理论^[13]。该理论针对TSSI直接将矛盾排除在外，并对数学描述不完备的问题进行改进，解决了以下问题：数学描述不完备问题（定义了6种真性距离与语义信息之间的关系）、语用偏好问题（距离公式中添加权重）、矛盾不可计算问题（引入介于语义真假的第三态）、与CSI度量不一致问题（距离定义中加入概率描述），并且引入了效用与决策理论来分析语义信息（距离函数中引入效用参数）。这6种距离与语义信息之间的运算关系分别为：

- 距离与语义关系1： $\Delta_{\min}(A, T) = \Delta_{\min}(w_a, w_T)$ ，其中 $w_a \in W_A$ 。
- 距离与语义关系2： $\Delta_{\max}(A, T) = \Delta_{\max}(w_a, w_T)$ ，其中 $w_a \in W_A$ 。
- 距离与语义关系3： $\Delta_{\text{sum}}(A, T) = \sum_{w_a \in W_A} \Delta(w_a, w_T) / \sum_{w_b \in B} \Delta(w_b, w_T)$ 。
- 距离与语义关系4： $\Delta_{\text{av}}(A, T) = \sum_{w_a \in W_A} \Delta_{\max}(w_a, w_T) / |W_A|$ 。
- 距离与语义关系5： $\Delta_{\min}^{\gamma}(A, T) = \gamma \Delta_{\min}(A, T) + (1 - \gamma) \Delta_{\max}(A, T)$ ，其中 $0 \leq \gamma \leq 1$ 表示权重。
- 距离与语义关系6： $\Delta_{\text{ms}}^{\lambda}(A, T) = \gamma \Delta_{\min}(A, T) + \lambda \Delta_{\text{sum}}(A, T)$ ，其中 $0 \leq \gamma \leq 1$ 与 $0 \leq \lambda \leq 1$ 表示权重。

在上述运算中， $\Delta(w_a, w_T)$ 表示命题 w_a 与真性命题 w_T 之间的真性距离， $\Delta(A, T)$ 表示命题 w_a 在真性命题 w_T 定义的场景 T 中具有的语义信息 A 。

D' ALFONSO结合CSIT与TSSI的思想，给出了一种新

的 $\Delta(w_a, w_T)$ ，其改进的语义信息公式为：

$$\text{val}(w) = \frac{t}{n \times 2^n} \quad (4),$$

其中， n 表示逻辑空间中逻辑命题的数量， t 表示命题 w 在实际情况中包含的所有真命题的数量。需要说明的是，公式 (4) 已经将真命题及其存在的场景整合在内了，因此该公式更倾向于描述命题 w 的性质，而非与真性命题的距离。

2 后香农时代的语义度量理论

事实上，上一节提及的语义度量理论均针对简单的逻辑命题，目前还没有一套通用的理论能够有效衡量人类语言中语义信息的衡量体系。但是，也有很多研究者从各个角度，对语义信息进行了尝试性描述，比如代数信息理论^[14-15]、通用语义信息理论^[16-17]、语义编码理论^[18]等。本节举出几个对语义通信影响较大的度量方式。

2.1 经典语义信息理论的扩展

对语义通信影响最大的是 BAO 等在 2011 年发表的研究^[19]。该研究首次提出语义通信系统中的本地知识库与共享知识库概念，仍以熵函数为核理论思路。BAO 等认为，扩展 CSIT 可以使该理论体系完备地描述语义信息。该理论同样对语义描述的适用范围做了严格假设。在该研究中，语义信源发出的消息被假定为真，这样就避免了 BCP。定义信源语义信息为一个元组 $\{W_s, K_s, I_s, M_s\}$ ，信宿语义信息也为一个元组 $\{W_r, K_r, I_r, M_r\}$ ：

- W_s, W_r ：表示信源/信宿可能观察到的世界模型。
- K_s, K_r ：表示信源/信宿的背景知识库。
- I_s, I_r ：表示信源/信宿的推理程序。
- M_s, M_r ：表示信源/信宿用来编码/解码消息的消息生成/解释器。

与 CSIT 不同之处在于，该研究重新定义了消息 x 的逻辑概率与语义信息：

$$m(x) = \frac{\mu(W_x)}{\mu(W)} = \frac{\sum_{w \in W, w \mapsto x} \mu(w)}{\sum_{w \in W} \mu(w)} \quad (5),$$

$$H_s(x) = -\log_2(m(x)) \quad (6),$$

其中， $\mu(W_x)$ 表示消息 x 在世界模型中对应模型集合的逻辑概率， \mapsto 表示一般命题可满足关系，即消息 x 的模型， $\mu(W)$ 表示世界模型的逻辑概率。需要注意的是，在 CSIT 的定义中，分子为命题 x 包含的所有原子命题数，分母为当前情况下的所有原子命题，这是两者的主要差异。

如果存在知识库 K ，那么上述语义信息量被定义为逻辑概率的条件信息：

$$m(x|K) = \frac{\sum_{w \in W, w \mapsto K, x} \mu(w)}{\sum_{w \in W, w \mapsto K} \mu(w)} \quad (7),$$

$$H_s(x|K) = -\log_2(m(x|K)) \quad (8)。$$

需要注意的是，上述条件熵是由逻辑概率函数定义的，其与经典统计中的概率不同，没有可加性，这也是我们认为该描述存在的最大问题。3.1 节将会解释没有可加性的量并不能被熵函数唯一表示。更进一步地，这里的逻辑概率没有给出具体的计算方式。而我们认为，逻辑概率与统计概率之间的映射关系才是求解语义信息量的核心问题。

2.2 其他语义度量理论

KOLCHINSKY 等^[20]将语义信息定义为“物理系统对其环境的信息，这种信息对于系统随时间维持自身存在是因果必需的”。该定义基于系统及其环境的内在动力学，同时借鉴了信息理论和非平衡统计物理学的思想，产生了信息度量的非负分解，并将其分为“有意义的比特”和“无意义的比特”。更进一步地，该定义为表达与“语义信息”相关的一系列概念提供了一个连贯的量化框架，如“信息价值”“语义内容”和“代理性”。

KOUNTOURIS 等^[21]则利用 AI 来定义语义信息，基于过程动力学、信号稀疏性、数据相关性和语义信息属性，将 AI 技术中的信息生成、信息重构与经典通信中的信息传输进行有机结合。他们的实验基于语义进行采样和通信，展示了重构误差和驱动误差成本，以及生成的无信息样本数量。

华为香港理论研究部同样希望借助 AI 来研究语义信息^[22-23]。但是由于 AI 本身是一个“黑盒子”，难以进行量化，语义信息和 AI 理论研究的过程是相辅相成的。他们认为，语义通信一定是语义问题和语用问题的结合，不能分开考虑。语义信息度量问题可能是 ANN 可存储信息量的问题，它采用 Korner 熵来描述 AI 状态数与语义信息量^[24]。在假定 ANN 模型没有推理能力的前提下，Korner 熵可以近似估计 ANN 模型参数与可存储状态数之间的关系。然而截至目前，该研究尚未取得实质进展。

3 语义度量的假设条件

基于上述语义度量理论的发展历史，本文结合当前已有的研究，对语义度量中的假设进行总结和分析，试图找到能够用来描述语义信息的相关量。

3.1 熵函数与语义信息

香农利用热力学中的熵公式度量信息量并非偶然^[1]，因为由符号表示的信息满足4条公理，即对称性、扩展性、极值性与可加性。符合这4个公理描述的信息量可以被度量为熵，且这种表示是唯一的——这被称为熵的唯一性定理。在熵唯一性定理证明中，可加性是该定理存在的必要条件，而其他3项则可以适当弱化。如下为4条公理与熵唯一性定理^[25]：

- 熵的对称性：对于任意 K , $H_K(\mathbf{P})$ 对 \mathbf{P} 的所有分量的连续和对称。
- 熵的扩展性：对于任意 K , 有 $H_{K+1}(P_1, P_2, \dots, P_K, 0) = H_K(P_1, P_2, \dots, P_K)$ 。
- 熵的极值性：对于任意 K , 有 $H_K(P_1, P_2, \dots, P_K) \leq H_K(\frac{1}{K}, \dots, \frac{1}{K})$ 。
- 熵的可加性：对于 $M = \sum_{i=1}^K m_i$, 存在：

$$H_M(P_1Q_{11}, P_1Q_{21}, \dots, P_1Q_{m_11}, P_2Q_{12}, P_2Q_{22}, \dots, P_2Q_{m_22}, \dots, P_KQ_{m_KK}) = H_K(P_1, P_2, \dots, P_K) + \sum_{k=1}^K P_k H_{m_k}(Q_{1k}, Q_{2k}, \dots, Q_{m_kk}) \quad (9)$$

- 熵唯一性定理：令 $H_K(P_1, P_2, \dots, P_K)$ 是概率矢量 $\mathbf{P} = (P_1, P_2, \dots, P_K)$ 的非负实函数，其中 $P_k \geq 0$, $\sum_{k=1}^K P_k = 1$ 。若 $H_K(P_1, P_2, \dots, P_K)$ 满足上述4条公理，则有 $H_K(\mathbf{P}) = -\lambda \sum_{k=1}^K P_k \log(P_k)$, 其中 λ 为一个正常数。

熵能否作为语义信息的唯一度量指标，取决于语义是否满足上述4条公理。

事实上，人类语言具有非平稳性、模糊性，语言之美也正体现在其模糊性中。正是得益于这种模糊性，人类才能欣赏各种文章、乐章等，甚至从中体会出一些“非语义”“非语法”的信息，也正是这种模糊性才使得语言并非简单的词汇叠加，而是在叠加过程中形成有结构、有组织的序列。这种构成使得该叠加产生了 $1+1>2$ 的效果，否则就会出现“断章取义”的问题。因此，语义需要依托语法编排才能存在。语义信息既不满足对称性，更不满足可加性。熵公式并不能作为语义信息的唯一度量形式。语义信息是一种更高维的信息存在形式。此外，可加性这一公理的本源来自于范畴论，若要进一步研究其内禀假设，还需要结合范畴论中的相关理论展开分析，这里不再赘述。

3.2 关于语义信息的假设

类似于熵的4条公理，在语义信息漫长的研究历程中人们也总结出了4条假设^[6-7,13]。如果将消息 m 中具有的语义信息量标记为 σ , 则有：

- 数据假设： σ 是基于一个或者一组数据 d 的。
- 格式假设：数据 d 是格式良好的。
- 意义假设：每一个数据 d 是具有意义的，且记为 δ_d 。
- 真性假设： σ 是具有真性的。

结合上述4条假设，基于CIT的经典通信要解决的问题是如何编码数据 d , 语义通信要解决的问题是如何编码 σ 。假设组成消息 m 的数据为 $\{d_i; i = 1, \dots, n\}$, σ 并非是每一个数据 d_i 的意义 δ_{d_i} 的简单叠加，而是在 d_i 变得格式良好的过程中，进一步消除不确定性，使得 σ 小于等于 δ_{d_i} 的直接叠加，因此类似于消息 m 为数据 d 的函数， σ 也是关于 δ_{d_i} 的一个函数：

$$m = wf_d(d) \quad (10)$$

$$\sigma = wf_{\delta}(d) \quad (11)$$

其中， $wf_d()$ 表示数据格式函数， $wf_{\delta}()$ 表示意义格式函数。

3.3 语义空间定义

为了更好地描述上述语义信息假设，本文进行如下定义：

- 数据 d ：人能够感受到的且可以记录世界状态变化的物理载体。
- 数据空间 $D = \{d\}$ ：某一种物理载体的集合。
- 语义空间 $\sigma = wf_D(\delta_D)$ ：某一种物理载体 D 所包含的意义 δ_D , 以及其上所有格式函数 $wf_D()$ 所产生的语义信息集合。

针对上述对语义信息的分析可以发现，语义信息是基于数据存在的。一句话的含义不会因是否有通信而发生变化，语义信息是数据的内禀属性。因此，在语义信息的研究中，不考虑信息产生方和接收方的信息，即认为通信双方是绝对理性客观的。类似地，由熵函数描述的数据信息，也是客观存在的，不会因是否存在通信而改变。

此外，也有多种物理载体 d 可以承载语义，比如汉语、英语、图片、视频等。然而，不同载体记录语义信息的能力是等效的。基于数字化的信息时代都建立在该假设之上。因为从数字的起源来看，其本质也是一种自然语言，同样被视为一种语义信息载体。试想，对于一个不懂英文的人，面对英文时看到的就是 26 个字母得到的编码，无法感知其语义含义；而懂英文的人由于知道英文的语法与词意，因此可以

解读其语义含义。对于最早的程序员而言，一些机器码是有具体含义的，因为他们了解机器码的语法。而对普通人而言，机器码可能就是一堆无意义的0和1组成的序列。我们可以将上述假设归纳为^[6,26]：

- 属性假设：意义是数据的内禀属性。
- 等价假设：不同物理载体承载语义信息的能力等价。

4 语义信息熵的定义与性质

语义信息不能使用熵公式进行唯一刻画，但或许可以借助这种形式确定其度量的上下界。基于数据假设，数据度量应该是语义信息度量的一种特殊形式。本文从该特殊形式出发，对理想中的极限情况进行定义。

4.1 理想语义空间与数据空间

考虑一个理想中的数据集 $D = \{d_i\}$ ，其内单个数据能够独立无关联地承载语义信息，并且承载的语义空间与其数据空间一一对应。此时该数据集描述的语义具有渐进均分性，基于熵公式计算的数据信息量，等于该数据集承载的语义信息量。此时 $wf_D()$ 与 $wf_\delta()$ 不产生作用，且格式假设不存在，那么理想数据集可以定义为：

- 元数据空间 D ：一种物理载体 $D = \{d_i\}$ ，其内单个数据能够无关联地承载语义信息，并且承载的语义空间 $\{\delta_D\}$ 与其数据空间 D 一一对应。其中，每一个 d_i 被称为一个元数据。
- 元语义空间 Δ ：元数据空间 D 所承载的语义空间 $\{\delta_D\}$ 。其中，每一个元数据 d_i 所代表的意义 δ_{d_i} 被称为一个元语义。

事实上，香农在定义通信模型的过程中，明确指出不考虑符号背后的含义，认为符号是独立同分布的，具有渐进均分性，是一种典型集，并且样本概率近似为总体概率，从而近似得到数据熵。该假设是一个很强的假设（此时格式假设不存在），在这种假设下符号独立，因此可以认为其承载的语义信息也是独立的。可以看出，香农通信模型中的信息量为语义信息的一种特殊情况，元数据空间的语义信息就属于这种情况。

但是，一旦考虑符号背后的含义，以及含义之间的关联和语法赋予的关联，符号就不再是独立同分布的，因此不再属于典型集，甚至不是线性的、平稳的、各态历经的。样本概率无法近似总体概率，数据熵无法描述语义信息，经典信源编码、信道编码定理不再适用。例如，选择不同的符号种类或者符号组合作为单个数据时，所选数据单元的概率分布会存在差异，最终计算得到的数据熵也不相同。

这是因为，数据本身具有含义，描述含义的物理载体具有一定格式。数据分布的改变反映了数据格式的改变。数据分布实际上包含了格式假设，使用统计概率代替了数据中一部分语义关联性。因此，语义信息与不同物理载体的数据分布相关。

此外，经典信息理论基于数字化和编码，这使得熵函数可以实现精确计算。但是如果考虑信息的原本描述，其往往是连续的，而非离散的。连续变量的熵由微分熵和绝对熵构成，其信息量趋于无穷，无法计算。近似计算需要对绝对熵进行离散处理，本质是对描述物理世界的数据进行重新采样。而语义编码一般基于 ANN 实现，具有概率意义，属于连续变量。因此，语义信息也与信源的采样方式或者数据描述物理世界的方式有关，这一般被称为语义采样。

图 2 为从数据空间转为语义空间的示意图。在图 2 (a) 所示的数据空间中，单个数据在经典信息理论中是独立同分布的。图 2 (b) 则进一步纳入数据的语义属性。意义相近的数据 1 与数据 2 属于同一个语义空间，针对其数据的编码距离也应更近。此时数据之间存在关联，不再是独立分布，无法用典型集描述。图 2 (c) 展示了数据到语义之间的非线性映射关系。由于数据存在多义、歧义，因此在意义上的关联是非线性的。真实物理载体的单个数据并非只具有意义，而是同样包含一定的关联信息。虽然本文认为考虑了语义后的数据集属于非典型集，无法直接用熵函数描述，但是符合典型集的数据为非典型集的一种极限情况。或许图 2 (c) 展示的语义信息可以基于极限情况，由图 2 (a) 中数据的熵函数进行逼近。

4.2 语义信息定义与特征

在实际情况中，一般物理载体的数据格式 $wf_D()$ 与语义格式 $wf_\delta()$ 并不等价，因此，其熵函数计算得到的信息量并不等于其语义信息量。按照 3.2 节中的四大假设，以及 4.1 节中的分析，本文将实际物理载体承载的语义信息分为两部分考虑。

对于实际数据集 $E = \{e_i; i = 1, 2, \dots, n\}$ ，其中 n 表示数据集 E 中具有单个数据的个数，单个数据 e_i 之间都具有语义关联性，此时熵函数无法描述语义。针对某消息 $M = \{e_j; j = m_1, m_2, \dots, m_k\} \subseteq E$ 由 E 中的 k 个单数据组成，那么定义消息 M 的语义信息量：

组成消息 M 的数据 $\{e_j\}$ 本身包含的语义信息量与消息 M 基于的实际数据 E 所包含的语义关联信息量，在语义空间中的总和可以表示为：

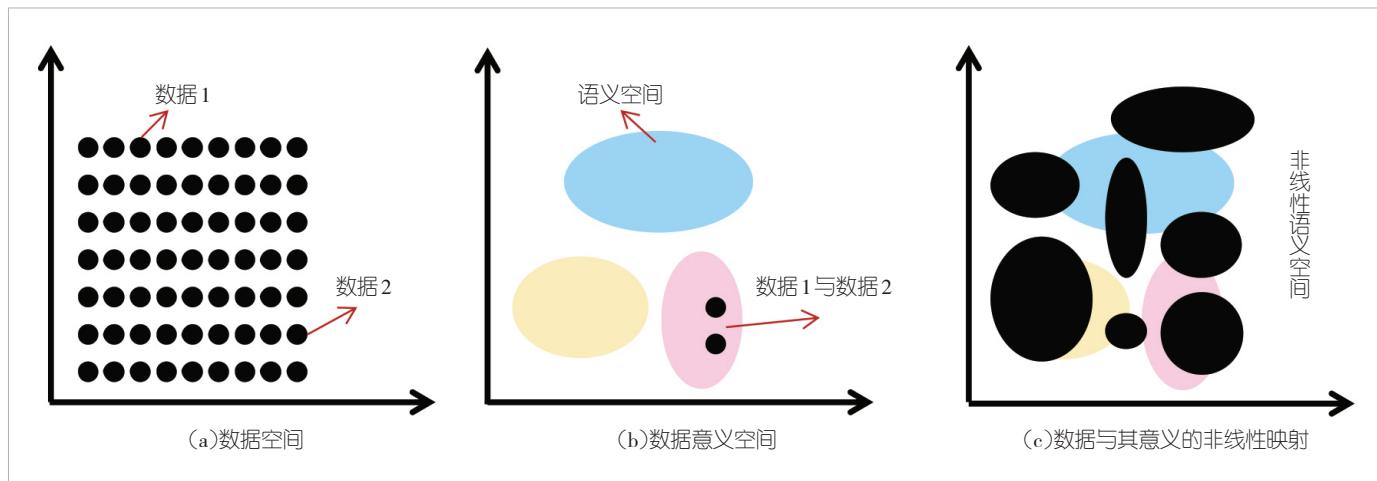


图2 数据空间与语义空间之间映射的非严格表示

$$S(M) = wf_{\delta}(\delta_M) = \lambda(\delta_M) \otimes cor(\delta_E) \quad (12),$$

其中, \otimes 表示高维运算, 是在不能确定语义格式函数 $wf_{\delta}(\cdot)$ 具体形式时的权宜表示。 $cor(\delta_E)$ 表示实际数据 E 所包含的所有单个数据之间的语义关联, 称为语义关联信息。 $\lambda(\delta_M)$ 表示组成消息 M 的单个数据意义的集合 $\{\delta_{e_j}; j = m_1, m_2, \dots, m_k\}$, 称为语义特征信息。

图3为语义信息的可视化示意图。其中, 蓝色表示物理世界的整体描述, 即本文定义的数据。横坐标表示数据的语义关联信息。由于语义关联信息包含了单个数据意义之间的关联, 以及语义赋予的确定性信息, 因此无法严格地示意出来。纵坐标表示数据的语义特征信息。元数据空间中的数据意义集合既等价于语义特征信息, 也等价于语义信息。然而, 实际数据空间中的数据, 其意义之间存在关联。

需要说明的是, 数据假设体现在语义特征信息里, 但同时实际数据的语义特征信息也包含了一部分格式假设, 因为

其存在一定的分布特征。而语义关联信息则是格式假设的另一大体现, 它不仅涵盖了分布特征, 还融入了真性假设。此外, 语义关联信息属于确定性信息, 其作用相当于消除语义特征信息中的一部分不确定性。语义信息的核心性质如下:

- 语义特征信息一定大于等于语义信息。

- 语义关联信息包括词义关联信息, 即词义与词义之间的关联。这一类信息类似于自然语言处理 (NLP) 中的词嵌入向量, 其与基于真性距离的 TSSI 相关。

- 语义关联信息包括语法关联信息, 即上下文之间的关联。这一类信息类似于 NLP 中的语言模型, 其与基于逻辑概率的 CSIT 相关。

根据四大假设, 语义信息包含数据本身的意义与意义编排。这里将数据本身的意义视为语义特征信息, 将意义编排产生的效果视为语义关联信息。根据4.1节中定义的语义空间, 这些关联或许无法通过离散变量完整描述, 可能需要采用连续变量描述。这就使得原本基于经典概率空间的熵函数, 扩展到基于概率分布函数的熵函数。其中, 微分熵或许可以用来表示上述语义特征信息, 绝对熵可以用来表示语义关联信息。

4.3 熵函数逼近下的语义信息

为了得到语义信息的界限, 需要进一步研究4个假设。

基于数据假设和格式假设, 可以得到:

- 定理一: 任意自然语言的数据熵下确界小于等于其携带的语义信息:

$$S(\delta_E) \geq \inf_{E \in \Omega} H(E) \quad (12),$$

其中, Ω 表示所有自然语言集合。

基于数据假设, 语义信息一定需要数据才能存在, 即数

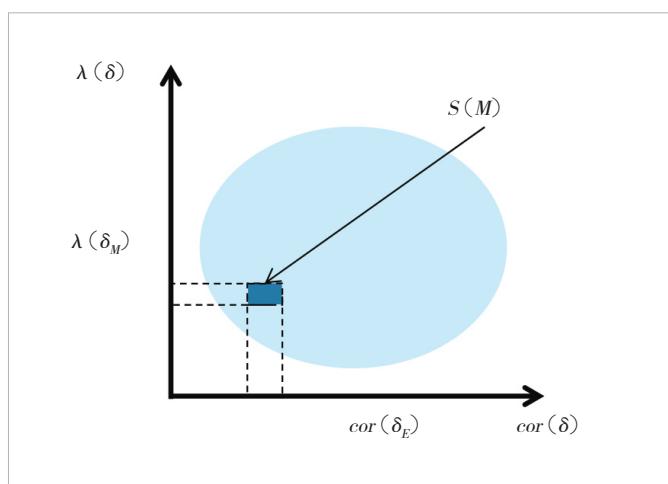


图3 语义信息的非严格示意图

据是语义存在的必要条件。因此定理一的成立可以直接证明。针对消息 M ，在元数据空间 D 中，其数据信息等于语义信息：

$$H_D(M) = S(M) \quad (13),$$

其中， $H_D(M)$ 表示采用物理载体 D 表达消息 M 熵的大小。但是对于真实数据 E ，其相比于元数据，由于存在格式假设，必然存在数据间关联。该关联体现在真实数据的不同分布形式中，即元数据分布必然比真实数据分布更均匀。这是由于关联关系消除了一部分不确定性。因此，真实数据熵下确界小于等于元数据熵：

$$H_E(M) \leq H_D(M) \quad (14).$$

推广到一般情况，可以得到数据熵下确界小于等于其携带的语义信息，即 $S(\delta_E) \geq \inf_{E \subset \Omega} H(E)$ 。

此外，基于意义假设，可以得到：

- 定理二：任意自然语言的数据熵上确界大于其携带的语义信息：

$$\sup_{E \subset \Omega} H(E) > S(\delta_E) \quad (15),$$

其中， Ω 表示所有自然语言集合。

基于意义假设，所有数据都具有一定的意义。针对消息 M ，在元语义空间 Δ 中，组成消息 M 的所有单个数据意义的集合，即该消息的语义特征信息等于语义信息，也等于：

$$\{\delta_d\}_{d \subset M \wedge d \subset D} = \lambda(\delta_M) = H_D(M) \quad (16),$$

其中， $\{\delta_d\}_{d \subset M \wedge d \subset D}$ 表示采用元数据空间 D 中消息 M 包含的所有单数据意义的集合。但是真实数据的语义信息小于等于其语义特征信息：

$$S_E(M) \leq \lambda(\delta_M) \quad (17).$$

推广到一般情况，可以得到数据熵的上确界大于等于其携带的语义信息。但是由于在语义空间中，格式假设一定存在，因此等于并不成立，因此 $\sup_{E \subset \Omega} H(E) > S(\delta_E)$ 。

根据上述两个定理，可以得到语义信息界为：

$$\sup_{E \subset \Omega} H(E) > S(\delta_E) \geq \inf_{F \subset \Omega} H(F) \quad (18).$$

其上界说明，数据熵描述的信息由于未考虑数据背后意义的关联性，均可以依靠语义信息进行进一步压缩。而下界说明，语义编码本质是找到一个最有效率的物理载体，该载体能够充分依靠语义关联性减少数据信息量，从而进行高效的语义描述。

更进一步地，熵的上确界与下确界之所以存在，本质是考虑了数据背后的含义以及含义之间的关联。对于经典信息

理论，数据属于典型集，样本概率可以近似为总体概率，因此数据熵能够收敛到一个确定的数据，并且这个数值可以用样本估计出来。但是，一旦考虑数据的语义，这种数据集不再是独立同分布，样本概率无法逼近总体概率，数据不再属于典型集，熵无法被准确估计，或者说此时估计的熵与所使用的数据集和数据采样方式均有关。因而，熵存在上确界与下确界。

针对消息 M ，语义编码需要找到合适的数据集与采样方式，尽可能有效地描述 M 的语义。此时的数据集与采样方式决定了语义关联信息，该信息为确定性信息，并被存储在知识库中。依靠知识库，语义通信可以进一步消除语义的不确定性，实现通信数据压缩，节省通信资源。根据定理一，其编码的最小长度为： $\inf_{F \subset \Delta} H(F)$ ，所能节省的最大数据量为 $\sup_{E \subset \Delta} H(E) - \inf_{F \subset \Delta} H(F)$ 。

5 结束语

本文从经典通信出发，详细介绍了其面临的主要困境，并指出语义通信可能是突破该困境的方式之一。而语义通信的研究由来已久，主要分为 3 条路，其中最重要也是发展最缓慢的是语义度量理论。在语义度量的漫长发展历史中，使用逻辑概率与真性距离描述语义信息是最重要的两类思想。此外，也不乏基于物理系统、基于 AI 等的语义信息度量理论。本文综合上述研究，提取出语义度量的 4 条基本假设，试图将 CIT 与 TSSI 等归纳入统一的语义度量理论中，实现概率视角下语义度量方式与真性距离视角下语义度量方式的有机结合，给出语义空间、语义信息等的具体定义。同时，基于熵函数，采用理想实验得到了语义信息的上下界，从而对语义编码进行实际指导。

致谢

感谢中兴通讯股份有限公司算法部胡留军、郁光辉、梁楚龙等专家对本研究的帮助！

参考文献

- [1] SHANNON C E. A mathematical theory of communication [J]. Bell system technical journal, 1948, 27(3): 379–423. DOI: 10.1002/j.1538-7305.1948.tb01338.x
- [2] WEAVER W. Recent contributions to the mathematical theory of communication [EB/OL]. [2025-11-05]. https://courses. ischool. berkeley.edu/i218/s15/Weaver_Recent-Contributions.pdf.
- [3] HORNIK K, STINCHCOMBE M, WHITE H. Multilayer feedforward networks are universal approximators [J]. Neural networks, 1989, 2(5): 359–366. DOI: 10.1016/0893-6080(89)90020-8
- [4] ITU. Framework and overall objectives of the future development of IMT for 2030 and beyond: ITU M.2160 [S]. 2023
- [5] BAR-HILLEL Y, CARNAP R. Semantic information [J]. The British

- journal for the philosophy of science, 1953, 4(14): 147–157. DOI: 10.1093/bjps/iv.14.147
- [6] FLORIDI L. Outline of a theory of strongly semantic information [J]. *Minds and machines*, 2004, 14(2): 197–221. DOI: 10.1023/B:MIND.0000021684.50925.c9
- [7] FLORIDI L. Is semantic information meaningful data? [J]. *Philosophy and phenomenological research*, 2005, 70(2): 351–370. DOI: 10.1111/j.1933-1592.2005.tb00531.x
- [8] TURING A. On computable numbers, with an application to the entscheidungsproblem (1936) [M]//The Essential Turing. Oxford University PressOxford, 2004: 58–90. DOI: 10.1093/oso/9780198250791.003.0005
- [9] RUMELHART D E, HINTON G E, WILLIAMS R J. Learning representations by back-propagating errors [J]. *Nature*, 1986, 323 (6088): 533–536. DOI: 10.1038/323533a0
- [10] LECUN Y, BENGIO Y, HINTON G. Deep learning [J]. *Nature*, 2015, 521(7553): 436–444. DOI: 10.1038/nature14539
- [11] ZHAO W X, ZHOU K, LI J, et al. A survey of large language models [EB/OL]. (2023–03–31) [2025–11–05]. <https://arxiv.org/abs/2303.18223>
- [12] BACCHUS F I. Representing and reasoning with probabilistic knowledge [M]. Cambridge: MIT Press, 1993
- [13] D' ALFONSO S. On quantifying semantic information [J]. *Information*, 2011, 2(1): 61–101. DOI: 10.3390/info2010061
- [14] KOHLAS J, SCHNEUWLY C. Information algebra [M]. Berlin: Springer–Verlag Berlin Heidelberg, 2009
- [15] LANGEL J. Logic and information: a unifying approach to semantic information theory [EB/OL]. [2025–11–05]. <https://docslib.org/doc/4260093/logic-and-information-a-unifying-approach-to-semantic-information-theory>
- [16] JUBA B, SUDAN M. Universal semantic communication I [EB/OL]. [2025–11–05]. <https://people.seas.harvard.edu/~madhusudan/papers/2007/juba-full.pdf>
- [17] JUBA B, SUDAN M. Universal semantic communication II: a theory of goal-oriented communication [EB/OL]. [2025–11–05]. <https://people.csail.mit.edu/madhu/papers/2008/usc2.pdf>
- [18] WILLEMS F M J, KALKER T. Semantic compaction, transmission, and compression codes [C]//Proceedings of International Symposium on Information Theory. IEEE, 2005: 214–218. DOI: 10.1109/ISIT.2005.1523325
- [19] BAO J, BASU P, DEAN M K, et al. Towards a theory of semantic communication [C]//Proceedings of IEEE Network Science Workshop. IEEE, 2011: 110–117. DOI: 10.1109/NSW.2011.6004632
- [20] KOLCHINSKY A, WOLPERT D H. Semantic information, autonomous agency and non-equilibrium statistical physics [J]. *Interface focus*, 2018, 8(6): 20180041. DOI: 10.1098/rsfs.2018.0041
- [21] KOUNTOURIS M, PAPPAS N. Semantics–empowered communication for networked intelligent systems [J]. *IEEE communications magazine*, 2021, 59(6): 96–102. DOI: 10.1109/MCOM.001.2000604
- [22] ORLITSKY A, ROCHE J R. Coding for computing [C]//Proceedings of IEEE 36th Annual Foundations of Computer Science. IEEE, 1995: 502–511. DOI: 10.1109/SFCS.1995.492580
- [23] YUAN D H, GUO T, BAI B, et al. Lossy computing with side information via multi-hypergraphs [C]//Proceedings of IEEE Information Theory Workshop (ITW). IEEE, 2022: 344–349. DOI: 10.1109/ITW54588.2022.9965914
- [24] HARANGI V, NIU X Y, BAI B. Generalizing Körner's graph entropy to graphons [J]. *European journal of combinatorics*, 2023, 114: 103779. DOI: 10.1016/j.ejc.2023.103779
- [25] FEINSTEIN A. A new basic theorem of information theory [J]. *Transactions of the IRE professional group on information theory*, 1954, 4(4): 2–22. DOI: 10.1109/TIT.1954.1057459
- [26] WU J. The beauty of mathematics in computer science [M]. Boca Raton, FL: Chapman and Hall/CRC, Taylor & Francis Group, 2019. DOI: 10.1201/9781315169491

作者简介



唐雪, 中兴通讯股份有限公司副总裁、无线及算力产品经营部战略架构总经理; 主要负责无线及算力核心产品全球战略规划、核心技术规划、高端市场建设等。



许进, 中兴通讯股份有限公司无线算法部部长; 主要从事新型调制编码、网络编码、语义通信等新技术的预研; 发表论文10余篇。



冯雨龙, 中兴通讯股份有限公司算法工程师; 主要研究领域为语义通信、人工智能、机器学习; 发表论文7篇。

智能超表面自动化部署解决方案



Automatic Deployment Solution for Reconfigurable Intelligent Surface

李萍/LI Ping, 叶峥嵘/YE Zhengzheng, 窦建武/DOU Jianwu

(中兴通讯股份有限公司, 中国 深圳 518057)
(ZTE Corporation, Shenzhen 518057, China)

DOI: 10.12142/ZTETJ.202506010

网络出版地址: <https://link.cnki.net/urlid/34.1228.TN.20251219.1011.006>

网络出版日期: 2025-12-19

收稿日期: 2024-11-16

摘要: 智能超表面 (RIS) 打破了传统无线通信仅能被动适配信道的局限, 能够显著优化无线通信系统的性能。介绍了无线网络中 RIS 部署的主要场景和准则, 提出一种可行的 RIS 自动化部署解决方案的流程, 包括信号源位置确定、RIS 可部署区域选定、最佳部署方位选择、最佳码本集生成和动态协同。通过数学建模, 详细介绍了自动化部署问题中关键 RIS 位置朝向的确定方法, 并搭建了一套数字孪生仿真系统。仿真结果表明, 所提方案可以有效地给出 RIS 部署方式和性能量化建议。

关键词: 智能超表面; 自动化部署; 位置; 朝向; 动态协同

Abstract: The reconfigurable intelligent surface (RIS) breaks the limitation that traditional wireless communication can only passively adapt to channels, and significantly optimizes the performance of wireless communication systems. In this paper, the main scenarios and guidelines for RIS deployment in wireless networks are discussed. Furthermore, the procedure of an automatic deployment solution for RIS installation is proposed, including determination of signal source location, selection of RIS-deployable areas, selection of optimal deployment location and orientation, generation of optimal codebook sets, and dynamic coordination. Through mathematical modeling, a detailed introduction is given to the determination method of the RIS position and orientation in the automatic deployment problem, and a digital twin simulation system is built. The simulation results show that the proposed solution can effectively provide recommendations for RIS deployment methods and performance quantification.

Keywords: reconfigurable intelligent surface; automatic deployment; position; deployment direction; dynamic coordination

引用格式: 李萍, 叶峥嵘, 窦建武. 智能超表面自动化部署解决方案 [J]. 中兴通讯技术, 2025, 31(6): 70–74. DOI: 10.12142/ZTETJ.202506010

Citation: LI P, YE Z Z, DOU J W. Automatic deployment solution for reconfigurable intelligent surface [J]. ZTE technology journal, 2025, 31(6): 70–74. DOI: 10.12142/ZTETJ.202506010

智能超表面 (RIS) 是一种可编程的二维人工电磁表面。它通过调控每个阵元的相位状态来控制出射波束, 使其聚焦到期望的方向或位置, 从而实现对电磁信道环境的控制^[1–6]。RIS 的出现使得无线通信技术突破了“只能被动适应信道”的局限, 可以“主动改变和控制信道”, 实现信道容量提升、增秩、补盲、补弱等效果。特别是对毫米波等高频段无线网络而言, RIS 可以降低网络部署成本, 有效提升网络覆盖, 是 6G 无线通信领域中一种新的无线网络通信范式和潜在关键技术^[7–9]。

RIS 的部署是工程应用面临的一大难题。如何根据实际场景决定 RIS 面板最佳的布放位置和倾角朝向, 并优化码本库, 使得 RIS 发挥能力切实解决覆盖问题, 是通信厂商和运营商尤其关注的重要课题。另外, 如何能够使 RIS 的部署过

程尽量自动化、智能化, 降低对部署工程人员专业能力的要求, 实现流程的简化, 也是部署方案所必须考虑的。本文提出了一种 RIS 自动化部署的解决方案, 包括部署的基本准则、自动化部署的流程仿真结果等, 为 RIS 的实际商用提供可行性参考。

1 RIS 部署场景和准则

通常而言, 需要使用 RIS 的场景分为两类: 一是无线覆盖性能本身较差的场景, 二是需要通过 RIS 提升原有通信性能的场景, 如图 1 所示。典型的场景包括:

- 1) 由于植被、建筑物、山体等遮挡或信号反射, 导致室外道路、广场等区域无线覆盖性能较差的场景;
- 2) 由于墙体、玻璃、车身、集装箱外壳等造成信号穿透损失较大, 导致室内、楼宇、隧道、地下通道等区域覆盖薄弱的场景;

基金项目: 上海市未来产业试验场“揭榜挂帅”项目 (2024)

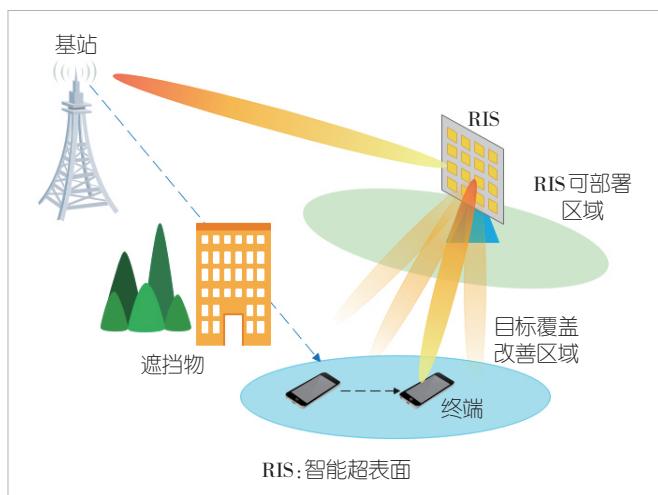


图1 智能超表面部署场景示意图

- 3) 需要构造直射径实现高精度定位的智慧物流、智慧医疗等场景;
- 4) 需要增加多径提升检测能力的目标检测、基站接收增强等场景;
- 5) 利用RIS进行高能效转发、增益或提升传输容量等场景^[10]。

为了尽量提升RIS部署后的效果, 主要的部署准则包括以下几方面:

1) 信号源与RIS需要有较强的直射径, 并且保证RIS拥有稳定的馈入功率。如果RIS板的入射信号没有较强的主径, 则代表馈入信号中存在多条功率相近的强径, 这些径大概率来自不同的入射角。对于特定码本来说, 出射信号会被散射到不同的方向, 导致入射功率无法被有效使用。这一现象会增加对邻区其他用户的干扰, 影响网络服务性能。

2) RIS与目标覆盖区域之间没有过多的遮挡物或者功率损耗。RIS面板的出射路径如果被遮挡, 信号将无法被目标覆盖位置的终端有效接收, 植被、建筑、行人、车体等遮挡都会大大增加路径损耗, 甚至导致信号传播方向偏离原始目标出射方向, 此时RIS对通信系统的增益效果就会大大降低。为了避免车辆行驶及行人往来对RIS面板的遮挡, 需考虑RIS合适的挂高及位置。

3) RIS的入射和出射角度在RIS面板的有效接收发射范围内。一般而言, RIS面板

都有额定的角度范围, 一旦超出, 将产生极大的插损并导致性能恶化。因此, 部署前需要明确所用RIS面板的基本信息。

2 RIS 自动化部署流程

对于一个需要部署RIS的场景, 从目标区域的选择到最后的网络部署调优, 主要的部署流程如图2所示。

2.1 确定信号源的位置和目标覆盖区域的范围

覆盖改善目标可以是一个固定点, 也可以是一片区域。根据RIS主要部署场景和准则, 在实际应用场景下可以事先确定信号源位置和目标覆盖区域的范围。在无线覆盖系统的诸多场景中, 需要支持对移动目标的实时跟踪, 并实现较大范围的覆盖。此时, 仅使用单个波束的静态或半静态RIS能力就会比较受限。由中兴通讯自主研发的动态RIS, 可以支持波束的实时切换, 实现广覆盖和目标跟踪的效果, 能够极大提升RIS所发挥能力的上限。

2.2 选定RIS的可部署区域范围

在实际工程部署中, 放置RIS面板的位置并不是随意选择的。例如, 我们不可能将其架设在道路中间阻碍人车通行, 也不可能将非透射板安装在窗前, 因其会遮挡光线和视野。很多时候RIS的安装位置需要结合具体部署场景综合研判, 并由运营商或通信厂商与物业、政企、商户等协商决定。比如, RIS可以安装在灯柱、墙角等位置。由于RIS轻巧灵活, 可被美饰为道路指引牌或者广告牌, 以减小RIS面板对周围环境的影响。因此, 在部署之前需要首先明确具体安装点位与适配形态, 以免重新规划。

2.3 选择RIS最佳部署位置和朝向

在选定可部署的区域范围以后, 我们就可以在其中选择最佳的位置朝向。最佳位置朝向的选择需要考虑目标覆盖区域的整体性能, 而不是单点的最佳性能。基于测绘、无人机航拍建模等方式, 我们可以确定基站、RIS可部署区域、目标覆盖区域的全局坐标系(GCS)。对RIS可部署区域进行网格化划分, 可以获得若干个RIS备选部署位置坐标, 编号为 $q=1,2,\dots,Q$ 。

在每个RIS部署位置下, 可以获得一组基站和目标覆盖区域相对RIS的局部坐标。假设RIS面板的水平阵子数、垂直阵子数分别为 M 和 N , 水平阵子间距 d_h , 垂直阵子间距

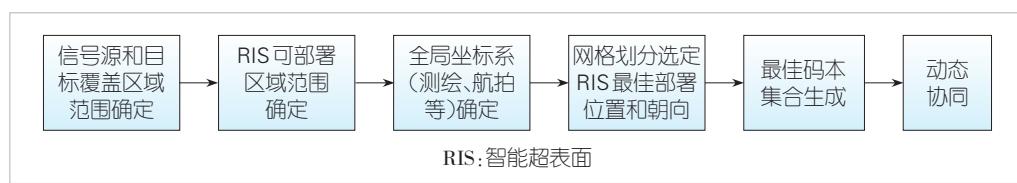


图2 RIS自动化部署流程

d_v ，信号波长为 λ ，RIS面板反射效率为 γ' ，RIS阵元功率增益为 G_e ，信号源发射功率为 P_t ，发射天线功率增益 G_t ，接收天线功率增益 G_r ，对于第 q 个RIS部署位置，基于RIS面板局部坐标系（LCS）的入射信号源球坐标为 $(r_{in}^q, \theta_{in}^q, \phi_{in}^q)$ 。目标覆盖区域通过网格化分割可以被间隔采样为 K 个采样点，相对RIS面板的球坐标表示为 $(r_1^q, \theta_1^q, \phi_1^q)$ ， $(r_2^q, \theta_2^q, \phi_2^q)$ ， \dots ， $(r_K^q, \theta_K^q, \phi_K^q)$ 。假设单元阵子为全向天线，对于第 k 个覆盖点的RIS功率增益可以近似表示为^[11]：

$$G_k^q = P_t \frac{\gamma' G_e G_t (\lambda M N)^2 d_h d_v \cos \theta_{in}^q \cos \theta_k^q}{64 \pi^2 (r_{in}^q r_k^q)^2} \quad (1)$$

通过设置一个满足用户需求的功率增益阈值 $G_{threshold}$ ，统计所有覆盖点中大于该阈值的数量：

$$n^q = \sum_k \delta(G_k^q - G_{threshold}) \quad (2)$$

其中， $\delta(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases}$ ，选择能最大化满足覆盖需求的位置。

若统计的满足功率增益阈值要求的点数相同，则进一步计算所有覆盖点的功率总和：

$$G_{total}^q = \sum_k G_k^q \quad (3)$$

这里我们选择功率总和最大的RIS位置作为实际的RIS部署位置。RIS面板朝向的建模可以通过方位角、下倾角和倾斜角来定义。通过调整全局坐标系向RIS面板局部坐标系的旋转量，可获得在不同面板朝向下入射信号源和出射目标覆盖区域的球坐标，从而选择出最佳的部署位置和朝向。假设全局坐标系下的 x 、 y 、 z 坐标为：

$$\hat{\rho} = \begin{bmatrix} x \\ y \\ z \end{bmatrix}^T = \begin{bmatrix} r \sin \theta \cos \phi \\ r \sin \theta \sin \phi \\ r \cos \theta \end{bmatrix}^T \quad (4)$$

其中， $[\cdot]^T$ 代表向量或矩阵的转置。全局坐标系变换到局部坐标系的旋转角度定义如下： z 轴固定 x 向 y 旋转的角度为 α ， y 轴固定 z 向 x 旋转的角度为 β ， x 轴固定 y 向 z 旋转的角度为 γ 。坐标系转换前后的示意图如图3所示。坐标系旋转矩阵可以表示为^[12]：

$$R = R_z(\alpha) R_y(\beta) R_x(\gamma) = \begin{bmatrix} +\cos \alpha & -\sin \alpha & 0 \\ +\sin \alpha & +\cos \alpha & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} +\cos \beta & 0 & +\sin \beta \\ 0 & 1 & 0 \\ -\sin \beta & 0 & +\cos \beta \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & +\cos \gamma & -\sin \gamma \\ 0 & +\sin \gamma & +\cos \gamma \end{bmatrix} \quad (5)$$

则局部坐标系下的 x 、 y 、 z 坐标可以表示为：

$$\hat{\rho}' = \hat{\rho} \cdot R \quad (6)$$

2.4 生成最佳码本集合

在RIS面板的位置和朝向选定完成以后，基站、RIS和目标覆盖区域的相对位置关系会被完全确定，此时可以生成用于指向目标覆盖区域不同点位的码本集合。RIS阵列水平垂直导向矢量表示为：

$$f_h = \left[1, \dots, \exp \left(-j2\pi \frac{md_h \sin \theta \sin \phi}{\lambda} \right), \dots, \exp \left(-j2\pi \frac{Md_h \sin \theta \sin \phi}{\lambda} \right) \right] \quad (7-1)$$

$$f_v = \left[1, \dots, \exp \left(-j2\pi \frac{nd_v \cos \theta}{\lambda} \right), \dots, \exp \left(-j2\pi \frac{Nd_v \cos \theta}{\lambda} \right) \right] \quad (7-2)$$

则对于码本 w ，RIS的三维空间方向图可以表示为：

$$PT(r, \theta, \phi) = G \times |w^H (f_h \otimes f_v)|^2 \quad (8)$$

其中， G 代表公式（1）所计算的功率增益。目标覆盖区域通过间隔采样分为 K 个采样点，相对RIS面板的球坐标表示为 (r_1, θ_1, ϕ_1) ， (r_2, θ_2, ϕ_2) ， \dots ， (r_K, θ_K, ϕ_K) ，对于第 k 个目标覆盖位置，最佳码本可以表示为：

$$w_k = \arg \max (PT(r_k, \theta_k, \phi_k)) \quad (9)$$

从而形成适用于目标覆盖区域的码本集合。

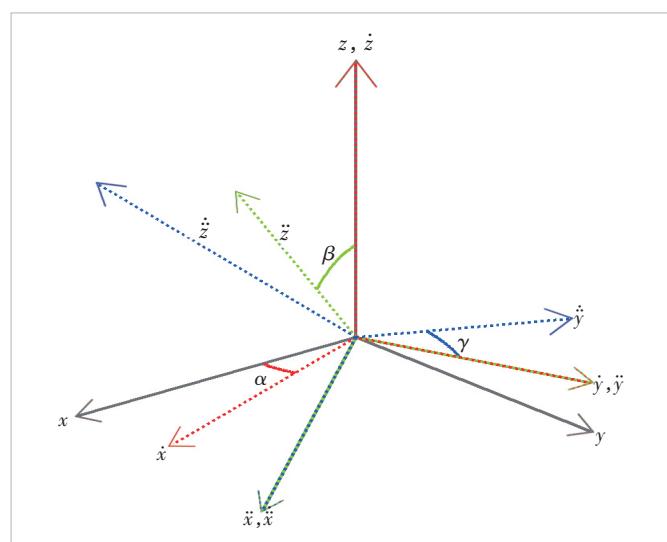


图3 局部坐标系与全局坐标系关系图

2.5 动态协同

使用静态或半静态方式的RIS无法实现波束的实时切换和用户设备(UE)不同位置的跟踪,对目标覆盖区域的性能提升十分有限。中兴通讯开创性研究的RIS动态协同方案与系统,可将不同覆盖点位实时切换到最佳码本,实现灵活跟踪和能效最大化,这对于新一代无线通信具有重要意义。对于2.4节生成的最佳码本集合 w_1, w_2, \dots, w_K ,使用动态协同功能可以实现实时切换,并在第 k 个目标覆盖点位自动将码本切换为 w_k ,实现最优覆盖。

3 仿真结果

基于上述的部署方案和理论建模,我们进行平台搭建和仿真,仿真条件和参数如表1所示。在26 GHz频段下,RIS水平垂直阵子数均为128,基站全局坐标系下的坐标位置为[150, 110, 20]。目标覆盖区域中,x坐标范围为80~130 m,y坐标范围为70~120 m,采样间隔均为2 m。RIS的3个备选部署位置坐标分别为[170, 90, 2.5]、[160, 100, 2.5]、[180, 80, 2.5],默认朝向为指向目标覆盖区域中心位置。在此基础上,仿真下倾角分别为0°、10°、-10°,水平角分别为0°、10°和-10°。目标覆盖区域范围、基站以及RIS可部署位置,在全局坐标系下的投影示意如图4所示。通过仿真,绘制在不同RIS部署位置、下倾角和方位角时目标覆盖区域内的功率情况累积分布函数(CDF)曲线,如图5所示,其中标记橙色圆圈的为RIS部署位置0,标记绿色方块的为部署位置1,标记红色三角形的为部署位置2,可以看到,当选择部署位置1,下倾角为-10°,方位角为-10°时的部署效果最好,因此我们最终选择该位置和朝向。

从仿真结果可以看出,设计的自动化部署和仿真方案可以有效给出RIS部署方式及性能量化建议。图6为数字孪生平台对中兴通讯上海研究所碧波湖外场进行三维地图建模及参考信号接收功率(RSRP)覆盖仿真结果,其中包含了对植被、建筑、毫米波基站等的建模,图中标注Ant1为基站位置。图7为使用RIS后RSRP的覆盖投影结果。对比图6和图7可以看出,使用RIS技术可以有效补偿毫米波覆盖空洞区域。

表1 自动化部署系统仿真参数

仿真条件和参数	说明
基站坐标	[150, 110, 20]
目标覆盖区域/m	X:[80:2:130], Y:[70:2:120]
频率/GHz	26
单子板阵子数	128×128
终端天线增益/dBi	23
面板法线朝向	指向目标覆盖区域中心为0°
智能超表面(RIS)位置	[170, 90, 2.5], [160, 100, 2.5], [180, 80, 2.5]
RIS入射的参考信号接收功率	-65 dBm, 相对RIS位置0, 其他位置按照基站和RIS的距离折算自由空间路损

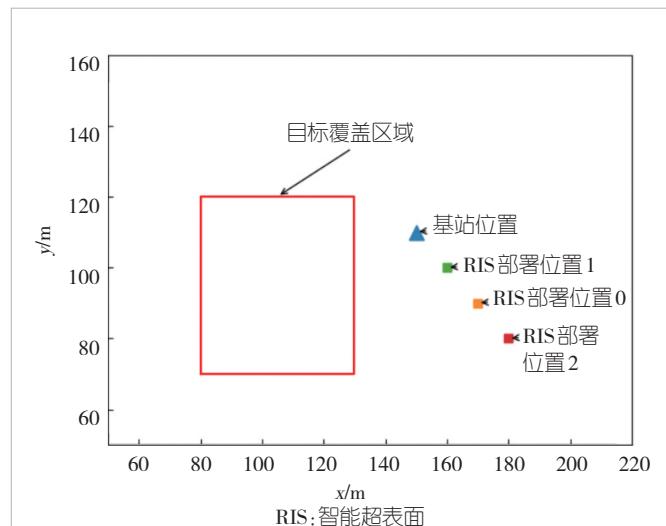


图4 自动化部署仿真全局坐标位置投影示意图

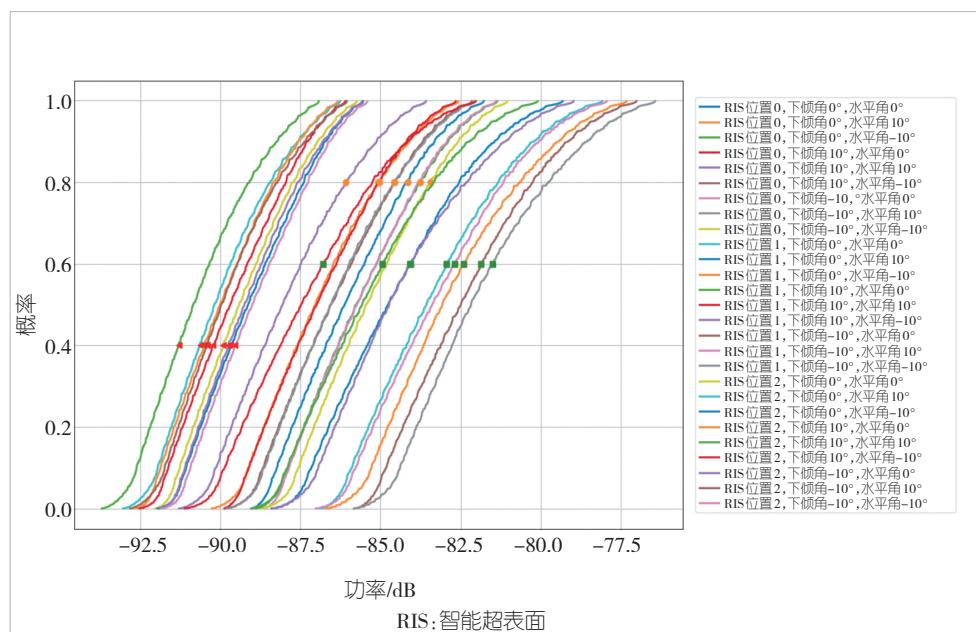


图5 不同RIS部署位置和朝向在覆盖区域内的接收功率累积分布函数(CDF)曲线

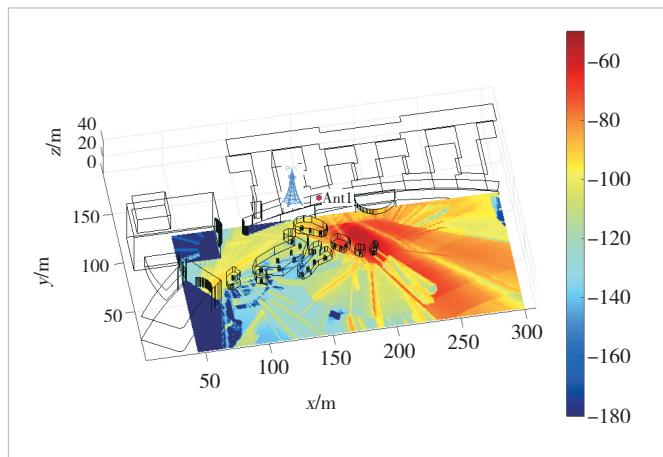


图6 没有使用智能超表面的参考信号接收功率覆盖图及三维建模

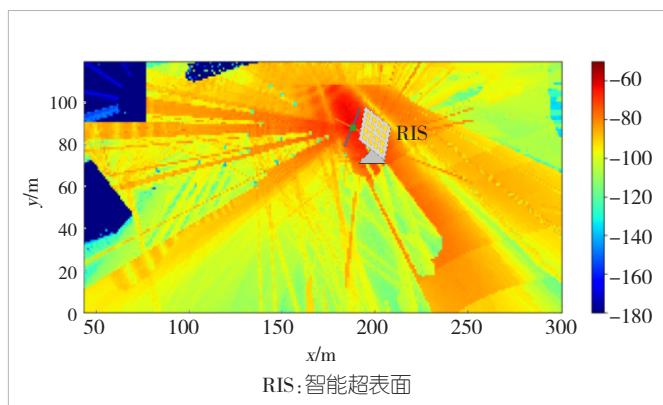


图7 使用智能超表面后的参考信号接收功率覆盖投影图

4 结束语

RIS的出现改变了传统无线通信只能被动适应信道的现状,它可以主动改变信道以提升信道容量,为未来无线通信技术的发展带来无限想象空间。如何进行有效部署是RIS走向落地应用的最基本问题。本文总结了RIS使用的场景和部署准则,提出了RIS部署流程并完成了相应的理论建模,同时搭建了基于该部署策略的数字孪生仿真平台。通过对实际部署场景的建模,筛选出了RIS最佳的部署位置和朝向,并基于此设计了码本集合。本文提出的方案为推进RIS产品化进程起到了重要作用。

参考文献

- [1] CUI T J, QI M Q, WAN X, et al. Coding metamaterials, digital metamaterials and programmable metamaterials [J]. *Light: science & applications*, 2014, 3(10): e218. DOI: 10.1038/lsa.2014.99
- [2] 赵亚军, 管梦楠. 6G智能超表面技术应用与挑战 [J]. 无线电通信技术, 2021, 47(6): 679–691. DOI: 10.3969/j.issn.1003-3114.2021.06.002
- [3] ABEYWICKRAMA S, ZHANG R, WU Q Q, et al. Intelligent reflecting surface: practical phase shift model and beamforming optimization [J]. *IEEE transactions on communications*, 2020, 68(9): 5849–5863. DOI: 10.1109/TCOMM.2020.3001125
- [4] WANG Y J, WEN D Z, MAO Y J, et al. RIS-assisted federated learning in multi-cell wireless networks [J]. *ZTE communications*, 2022, 21(1): 25–37. DOI: 10.12142/ZTECOM.202201004
- [5] 王子昂, 桑健, 李潇, 等. 室内热点场景多频段RIS辅助MIMO通信信道测量与建模 [J]. 中兴通讯技术, 2024, 30(3): 43–51. DOI: 10.12142/ZTETJ.202403008
- [6] 赵亚军, 郁光辉, 徐汉青. 6G移动通信网络: 愿景、挑战与关键技术 [J]. 中国科学: 信息科学, 2019, 49(8): 936–987
- [7] 崔亦军, 李萍, 叶峥嵘, 等. 无线网络中智能超表面部署研究 [J]. 无线电通信技术, 2022, 48(2): 291–296
- [8] ZHU Y T, XU Z Y, ZHANG H. T. Cooperative distributed beamforming design for multi-RIS aided cell-free systems [J]. *ZTE communications*, 2024, 22(2): 99–106. DOI: 10.12142/ZTECOM.202402013
- [9] ZHAO Y. Q, KE H Q, XU W, et al. RIS-assisted cell-free MIMO: a survey [J]. *ZTE communications*, 2024, 22(1): 77–86. DOI: 10.12142/ZTECOM.202401009
- [10] IMT-2030(6G)推进组. 6G总体愿景与潜在关键技术白皮书 [R]. 2021
- [11] TANG W K, CHEN M Z, CHEN X Y, et al. Wireless communications with reconfigurable intelligent surface: path loss modeling and experimental measurement [J]. *IEEE transactions on wireless communications*, 2020, 20(1): 421–439. DOI: 10.1109/TWC.2020.3024887
- [12] 3GPP. Study on channel model for frequencies from 0.5 to 100 GHz (Release 16): 3GPP TS 38.901 [S]. 2020

作者简介



李萍, 中兴通讯股份有限公司资深算法高级系统工程师; 主要研究方向为移动通信数字信号处理、多天线AI智能系统、通感一体化系统、智能超表面技术等; 负责多项国家级重大课题, 发表论文10余篇, 申请专利近100项。



叶峥嵘, 中兴通讯股份有限公司算法无线高级系统工程师; 主要研究方向为阵列信号处理、毫米波通信技术、智能电磁表面、5G/6G无线信道建模、基于AI的通信系统优化等; 参与多项国家级重大课题, 发表论文4篇, 申请专利8项。



窦建武, 中兴通讯股份有限公司正高级工程师, 中国电子学会电波传播分会委员, IMT-2030(6G)信道测量与建模组副组长; 主要研究方向为5G/6G无线信道建模、无人机通信、卫星通信、太赫兹通信、智能电磁表面等; 参与3GPP 5GCM/UAV/NTN等标准化工作, 所主导的Map-based Hybrid信道模型被3GPP及ITU国际标准化组织采纳, 主持并参与多项国家重大科技专项, 曾获中国专利金奖、上海市科技进步奖、中国通信学会科学技术奖一等奖等。



面向6G的天地一体化网络安全技术

Network Security Technology for Space-Ground Integrated Networks Towards 6G

陆海涛/LU Haitao^{1,2}, 周强/ZHOU Qiang¹,
代九龙/DAI Jiulong^{1,2}, 卢帆/LU Fan¹, 李锐/LI Rui^{1,2}
(1. 中兴通讯股份有限公司, 中国 深圳 518057;
2. 深圳市5G接入网安全技术研究及应用重点实验室, 中国 深圳 518055)
(1. ZTE Corporation, Shenzhen 518057, China;
2. Shenzhen Key Laboratory of 5G RAN Security Technology Research and Application, Shenzhen 518055, China)

DOI: 10.12142/ZTETJ.202506011
网络出版地址: <https://link.cnki.net/urlid/34.1228.TN.20240912.1755.002>
网络出版日期: 2024-09-13
收稿日期: 2024-03-25

摘要: 天地一体化是6G的基本组网架构,现有网络安全技术已不再适用新的安全风险与场景要求。分析了天地一体化网络广覆盖、海量接入、低时延、大带宽通信场景的安全需求。针对天地一体化网络面临的网络拓扑动态变化和卫星平台资源受限的挑战,重点分析了接入认证、星地回传安全传输、星间链路安全传输、轻量化的无线物理层安全、抗量子密码,以及量子保密通信和基于区块链的星地可信联盟等网络安全技术。认为未来有必要研究轻量化的接入认证、星地可信联盟等新技术,以解决当前天地一体化网络的安全与资源矛盾。

关键词: 星载基站; 星地协同; 动态拓扑; 轻量化; 量子通信

Abstract: The Space-Ground Integrated Network is the fundamental networking architecture for 6G, and existing cybersecurity technologies are no longer adequate to address the new security risks and meet the scenario requirements. This paper analyzes the security requirements for the wide coverage, massive access, low latency, and large bandwidth communication scenarios of the Space-Ground Integrated Networks, and focuses on network security technologies such as access authentication, satellite-ground backhaul security transmission, inter-satellite link security transmission, lightweight wireless physical layer security, quantum-resistant cryptography and quantum secure communication, and blockchain-based satellite-ground trusted alliance. It is believed that future research should prioritize the development of new technologies like lightweight access authentication and satellite-ground trusted alliances to resolve the conflict between security and resource constraints in the Space-Ground Integrated Networks.

Keywords: satellite base station; satellite-ground collaboration; dynamic topology; lightweight; quantum communication

引用格式: 陆海涛, 周强, 代九龙, 等. 面向6G的天地一体化网络安全技术 [J]. 中兴通讯技术, 2025, 31(6): 75-83. DOI: 10.12142/ZTETJ.202506011

Citation: LU H T, ZHOU Q, DAI J L, et al. Network security technology for space-ground integrated networks towards 6G [J]. ZTE technology journal, 2025, 31(6): 75-83. DOI: 10.12142/ZTETJ.202506011

经“十三五”以来的建设,中国已建成全球最大规模5G网络。截至2024年6月30日,中国累计建成5G基站383.7万座,占全球比例超过60%,下一步将统筹推进并加快信息通信业的高质量发展。“十四五”规划和2035年远景目标提出要建设高速泛在、天地一体、集成互联和安全高效的信息基础设施。天地一体是指将卫星通信网络作为地面

通信网络的重要补充与延伸,通过深度融合天基与地基网络资源,构建一个覆盖全球、无处不在、高速智能和安全可靠的通信网络^[1-4]。非地面网络(NTN)是天地融合通信的主流技术,第3代合作伙伴计划(3GPP)最早从R15阶段就开展了NTN技术研究和标准化工作,定义了NTN部署场景和相关网络架构;在R16阶段开展了NTN系统性解决方案的技术研究;在R17阶段基于NTN透明载荷的网络架构,针对卫星通信场景的多普勒频偏较大、通信端到端时延较大以及长距离传输带来的信号大幅度衰减等问题,设计了空口增

基金项目: 广东省重点领域研发计划项目(2020B0101120003)

强协议，引入了多种增强技术；在R18阶段以针对NR NTN的增强功能、进一步完善5G卫星组网能力为目标，主要包括支持10 GHz以上频段部署、覆盖增强、移动性和服务连续性增强，以及星上本地数据交换技术^[5]。2023年12月开始的R19阶段进一步增强NTN卫星通信网络能力，重点支持全基站上星的再生架构，扩展应用场景，开展下行覆盖增强、上行容量/吞吐量增强研究；R20及后续阶段将结合5G NTN遗留问题，针对多频段管理、高低轨卫星协同、核心网能力增强以及星地频谱共享等方面开展进一步研究。

3GPP R20已于2025年启动，将重点推进6G及其架构的初步研究。R21预计在2026年启动，将成为首个全球6G标准规范。与5G相比，6G网络需要应对超高速率、超低时延、超大规模连接等场景带来的新安全挑战。尤其是天地一体化网络引入的新安全风险与场景要求，现有网络安全技术不完全适用，这将极大影响6G网络研究与标准进程。因此，开展天地一体化网络安全技术研究十分必要。

本文首先介绍天地一体化网络架构的透明转发和可再生两种模式，然后分析天地一体化网络的广覆盖、海量接入、低时延、大带宽通信场景的安全需求，最后针对天地一体化网络面临的网络拓扑动态变化和卫星平台资源受限的挑战，分析接入认证、星地回传安全传输、星间链路安全传输、轻量化无线物理层安全、抗量子密码，以及量子保密通信和基

于区块链的星地可信联盟等网络安全技术。

1 天地一体化网络架构

天地一体化是指天基和地基网络单元能无缝协同工作，为用户提供连续覆盖和无感知的超高速、超低时延和超大规模连接性能。天地一体化网络由用户段、地面段和空间段组成。其中，用户段包括手机、车载和船载传感器、无人机通信模块、物联网终端等设备；地面段由地面基站和核心网组成，可实现地面网络通信；空间段由高中低轨卫星组成，根据承载设备类型可分为中继卫星、星载基站和星载核心网，构成天基网络，与地面网络协同工作，实现天地一体化通信。

天地一体化网络根据工作模式可分为透明转发模式和可再生模式，如图1所示。

透明转发模式：卫星被作为中继节点，透明转发终端和地面基站之间的射频信号，适用于海洋、沙漠、山地等地面网络难以覆盖的场景，可作为地面网络的覆盖补盲，提供通信连接和连续业务服务。其缺点是空口时延较大，难以支持高速率业务，且覆盖范围高度依赖于地面信关站的部署，不利于实现全球广域覆盖。

可再生模式：将地面网络的部分或全部网元部署于卫星上，相当于实现地面通信网的立体化，形成星地协同与高中低轨协同的天地一体化网络架构。根据星载网元的部署方

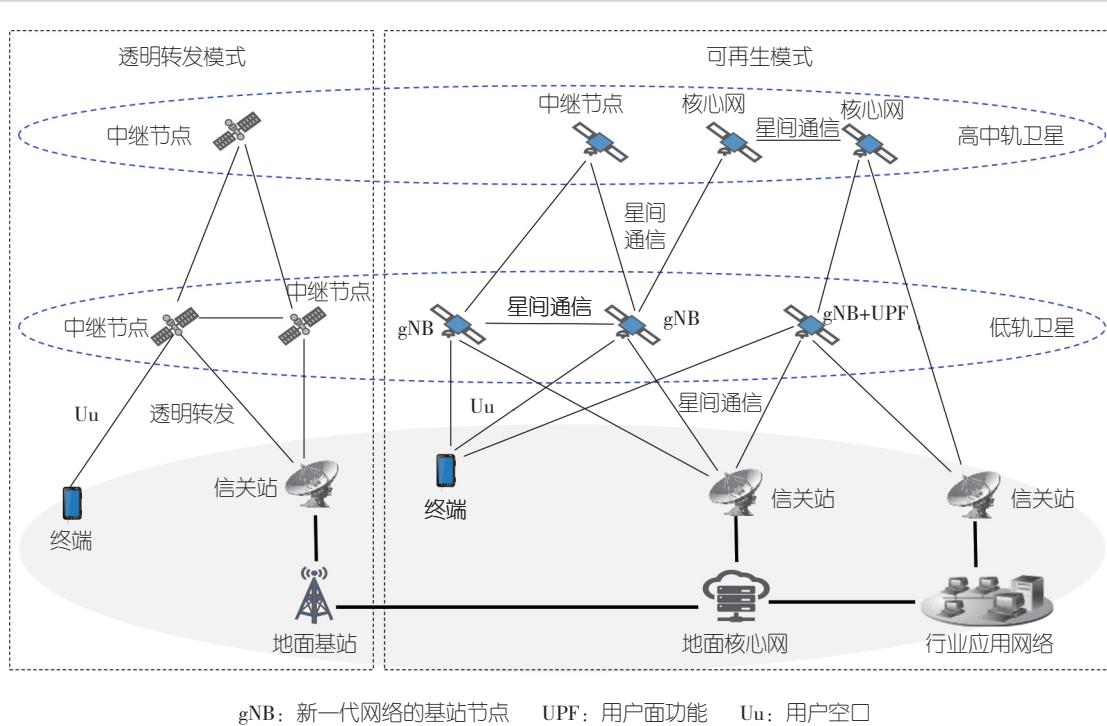


图1 天地一体化组网架构

式, 可再生网络主要可分为3种组网形式。

1) 星载基站+地面核心网: 基站上星部署, 卫星提供接入网功能。星载基站通过3GPP用户空口 (Uu) 连接地面终端, 同时星地回传通过卫星馈线链路连接地面信关站和地面核心网, 支持3GPP下一代接口 (NG)。星载基站部署在低轨卫星, 离地高度一般为150~2 000 km, 绕地球一圈只需1.5~2 h。因此, 星载基站的切换一直在持续进行。除了星载基站与终端之间的切换外, 星载基站与地面信关站和核心网的网络拓扑也一直在动态变化。此外, 星载基站是多颗部署, 星载基站之间星间链路通过激光通信承载3GPP基站互联接口 (Xn) 来进行互联。如果星载基站的轨道高度相同, 则星载基站之间的相对位置无变化, 网络拓扑可以保持相对稳定; 如果星载基站的轨道高度不同, 则星载基站之间的相对位置将持续变化, 网络拓扑也会一直变化。

2) 星载基站+用户面功能 (UPF) +地面核心网: 除了基站上星部署, 核心网的数据处理单元 UPF 也可以下沉并和基站一起部署在卫星上, 以减少传输时延, 实现低时延和大带宽的数据处理, 支持本地数据分流, 缓解核心网的数据传输压力。

3) 星载基站+星载核心网: 基站与核心网同时上星部署, 并且分布在不同轨道的卫星上。星载基站部署在低轨卫星, 支持低时延和大带宽数据处理。星载核心网部署在高轨卫星, 最大价值是实现全球覆盖的应急网络, 其卫星轨道高度约36 000 km, 保持与地面相对静止。理论上, 3个星载核心网即可实现全球覆盖, 这对于中国“一带一路”的通信保障具有非常重要的意义。星载基站与星载核心网之间的星间链路通过激光通信承载3GPP NG 接口来进行互联。由于星载基站的相对位置在不断变化, 星载基站与星载核心网的网络拓扑也一直在动态变化。此外, 星载基站和星载核心网通过地面信关站与地面核心网和应用网络连接, 支持星地协同, 实现天地一体化的统一网络。

上述是可再生网络的几个基本组网形式, 未来天地一体化网络的实际部署可能更复杂。例如, 星载基站可划分为分布式单元 (DU) 和集中单元 (CU), 二者分别部署在不同的卫星上, 甚至不同的轨道平面中。其中, 星载DU支持更高效的数据处理, 星载CU支持云化的卫星互联网。而星载核心网包含接入和移动管理功能 (AMF)、会话管理功能 (SMF)、策略控制功能 (PCF)、统一数据管理功能 (UDM)、认证服务器功能 (AUSF)、UPF和边缘计算云平台 (MEC) 等网络功能, 需要考虑哪些网络功能上星, 哪些保持在地面。

可再生模式是6G天地一体化网络的主要网络架构, 具有低时延、高带宽、广覆盖和灵活组网的特点。但相比于地

面网络, 网络安全面临的主要挑战有两点: 一是网络拓扑一直在动态变化, 需要持续地对终端和网元进行身份认证和连接控制; 二是卫星平台资源受限, 星载基站和星载核心网难以部署与地面网络相同的全面功能, 部分功能需要进行裁剪, 传统的安全技术和协议可能不适用于星上部署, 因此需要考虑轻量化的安全技术和协议。对于天地一体化网络面临的安全挑战, 接下来将分析具体的安全需求和安全技术。

2 天地一体化网络安全需求分析

天地一体化网络透明转发模式的网络结构比较单一, 其更多是作为一种过渡性方案。未来6G天地一体化网络将以可再生模式为主, 因此接下来的安全分析将基于该模式开展, 将主要结合天地一体化网络的特征和场景进行安全需求分析。

天地一体化广覆盖场景安全需求: 6G将融合地面基站、地面核心网、地面信关站、高中低轨星载基站、星载核心网及中继卫星等各类网络节点与设施, 实现天地网络融合及全球无缝覆盖。由于卫星网络节点位置在不断变化, 因此网络拓扑结构也在持续动态变化, 需要保障网络拓扑变化过程中网络节点的身份认证, 以及终端接入网络时的身份认证, 同时, 在多运营商融合组网场景下, 还需要满足多方身份互信的需求。

天地一体化海量接入场景安全需求: 6G天地一体化网络覆盖广, 连接终端数量多, 尤其是卫星物联网场景的海量终端接入, 需要考虑星载基站和核心网设备对认证消息的处理能力, 以支撑轻量化的认证技术。这包括简化认证流程、支持分组认证, 同时需要研究支持无线物理层认证等新认证技术。在保证安全性的基础上, 这些技术既可以满足海量接入认证需求, 又可以降低资源占用与成本开销。在智能家居、智慧城市等领域, 天地一体化网络需要连接大量的设备和传感器。这些设备和传感器需要进行身份验证和授权管理, 以确保它们只能访问其被授权的数据和服务。

天地一体化低时延、大带宽通信场景安全需求: 6G天地一体化网络受卫星轨道高度、频谱资源等因素的限制, 难以提供低时延的服务, 需要考虑简化接入协议来降低通信时延, 包括采用多用户共享接入 (MUSA) 技术。因此, 身份认证需要支撑简化接入信令交互流程时的认证机制, 满足单条信令即可实现的用户认证鉴权需求。例如, 在金融交易、实时通信等领域, 快速的身份验证和授权管理是必不可少的。此外, 6G通过采用更大规模天线、更大带宽等新空口技术, 大幅提升了空口传输速率。而传统基于计算复杂度的数据加解密技术可能存在性能瓶颈, 需要引入轻量化的数据加解密技术, 基于安全和通信的深度融合, 在保障安全性的

同时,提高系统的传输性能,满足6G网络高效的数据安全传输和高安全服务需求。

此外,天地一体化网络是多层次、立体化的网络架构,终端与网络之间的距离长且都处于开放的天空,更容易受到信号干扰、数据窃听和数据篡改攻击,因此需要研究星地空口安全保护方案,保障数据的机密性和信令的完整性,抵御星地空口分布式拒绝服务(DDoS)攻击和信号干扰攻击。天地一体化网络是全球覆盖的网络,由于外太空间全球开放,不同国家和不同运营商都可以部署星载通信节点,因此也面临着对星载设备的干扰、窃听与篡改威胁,需要探索伪星载基站探测和防御安全方案,研究抵御UE降级攻击、Xn切换降级攻击、射频干扰攻击、指纹分析攻击及位置追踪攻击等网络安全技术。

3 天地一体化网络的安全技术分析

目前业界对6G网络安全技术已有不少分析,普遍共识是传统“外挂式”“补丁式”的安全机制将向6G内生安全方向演进。IMT-2030(6G)推进组把内生安全列为6G潜在的十大关键技术之一,同时定义了内生安全的主动免疫、弹性自治、虚拟共生和泛在协同四大特征^[6]。文献[7]指出,轻量化

也是6G内生安全的主要特征之一,传统的接入安全技术无法满足海量设备连接和实时传输需求,需要引入轻量化的安全机制。针对6G天地一体化网络面临的网络拓扑动态变化和卫星平台资源受限的挑战,本文将重点分析以下网络安全技术。

文献[8]基于3GPP 5G安全域模型^[9],将6G网络的安全需求融入,尤其是天地一体化网络的安全需求,并提出6G网络的安全域模型,如图2所示。

相比于5G,6G网络安全域模型的最大特征是融入了卫星网络,新增了卫星接入网、卫星核心网和地面站等网元,在接入域、网络域和用户域进行安全增强。这主要涉及接入认证、星地回传安全传输、星间链路安全传输、轻量化物理层安全、抗量子密码迁移和区块链应用等网络安全技术。

1) 接入认证技术

天地一体化网络的星载通信节点位置在持续动态变化,并且处于开放的长距离空间环境中,更容易面临仿冒和劫持攻击风险,对此,系统首先需提供双向认证机制,基于共享密钥以及多层级密钥推演机制,保证终端接入网络的身份合法,以及被接入的网络的合法性。其次,对于行业专网应用,提供对接入企业专网的终端进行二次认证——该认证基于3GPP标准定义的二次认证,同时提供更细粒度终端访问控制

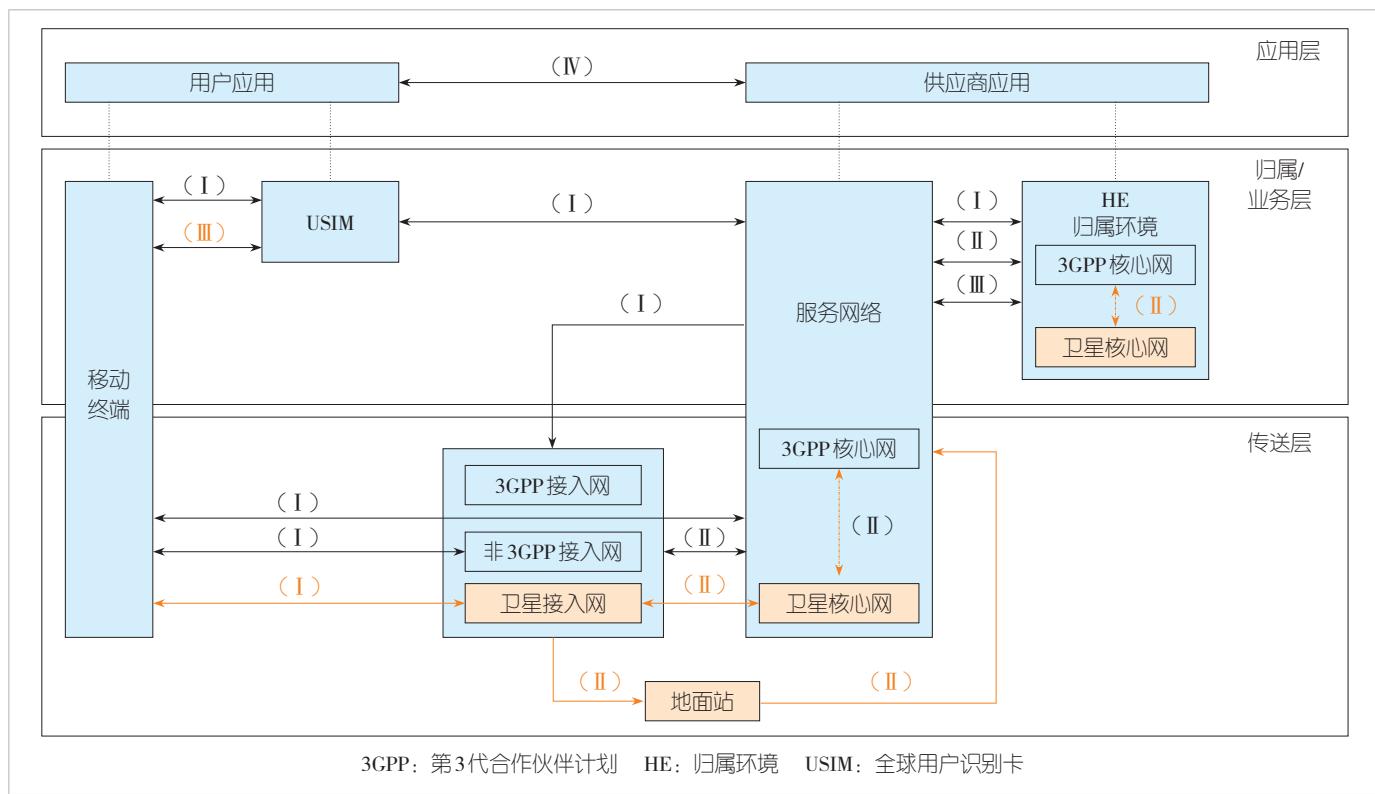


图2 6G网络的安全域模型

(包括根据终端位置、机卡一致性、终端属性等多因素进行安全认证), 以满足垂直行业业务对安全保密的更高要求。另外, 对于快速移动的星载基站引起的切换场景, 需提供安全的移动性管理; 涉及密钥变更等过程时, 要保持安全上下文和承载上下文的连续性。同时, 为保证天地一体化网络的高可用性, 建议提供星载基站的孤站自治功能: 如果与天基核心网或地面核心网的链路断开, 星载基站将启动断链保持模式, 确保存量终端的正常业务和认证不受影响, 同时支持新用户的接入认证, 最终实现业务不中断、安全不失效。

由于天地一体化网络覆盖广且通信距离长, 网络中存在海量终端的随机接入, 因此面临着并发认证数多和认证链路长的风险。而星载通信节点受限于硬件资源条件, 相比于地面网元更容易引起认证阻塞, 因此需要研究提供轻量化的认证技术。具体包括两个方面: 一是支持分组认证来降低并发认证数量, 例如将同一区域或相同用户的物联网感应器标识为一终端群组, 进行统一接入认证; 当终端群组的一个用户通过接入认证, 则该群组的所有用户都默认获得接入许可; 反之, 若一个用户被取消了接入许可, 则该群组的所有用户都被禁止接入网络^[10]。二是简化认证协议流程以缩短认证链路长度, 通过一条信令即可完成接入和认证。例如, 对第一条接入消息直接进行安全保护, 这种方式可以保护海量终端的空口接入消息, 防止攻击者窃听和篡改消息; 同时用户标识 (ID) 也进行加密随消息发送, 使得网络可以认证终端的合法性, 防止非法终端连接网络^[10]。

此外, 数字身份技术基于身份与访问控制、生物特征识别、机器学习模型等技术, 来增强访问安全和数据隐私, 在未来天地一体化网络接入场景中, 能够使用户和服务提供者之间建立起更加可靠的数字信任关系。具体方式包括: 扩展现有手机用户识别卡 (SIM) 作为数字身份的认证基础, 逐步增加安全芯片、数字证书、国家算法等能力, 完善认证/核验平台、卡端接入平台及业务合作平台, 集成面向个人、企业、政府的多场景应用, 搭建并完善可信数字身份体系^[11]。在天地一体化网络应用场景中, 需要处理大量的设备和传感器。这些设备和传感器使用数字身份技术进行身份验证和授权管理, 通过为每个设备分配一个唯一的数字身份, 实现设备的权限和访问级别的

管理与控制。

2) 星地回传安全传输技术

星地协同是6G天地一体化网络的一个主要特征, 它使用统一的网络架构和标准体制, 一体化的无线接入、传输和网络技术, 以及一体化的星地协同无线资源分配与业务管理机制, 为用户提供全域覆盖, 满足用户随时随地的通信需求^[12]。星地回传技术是星地协同的关键技术, 它能够使星载基站的数据通过馈线链路传输至地面信关站, 再由信关站转发至地面核心网和网管系统。星地回传具有4个突出特点: 一是星载基站与地面信关站之间进行单点通信; 二是通信数据量大, 满足基站与核心网间NG接口的回传数据带宽要求; 三是通信距离远、高速移动; 四是要根据卫星运动轨迹, 保证回传链路在不同信关站间的实时切换。

星地回传主要用于实现星载基站与地面核心网的NG接口连接。与地面NG接口有线传输相比, 星地回传增加了星地馈线链路的无线传输。现有地面NG接口使用互联网安全协议 (IPSec) 来保护NG接口安全, 因此星地回传的安全传输也以IPSec协议为主, 具体分为基站IPSec和信关站IPSec两种方式。

基站IPSec是指星载基站从信关站动态获得公网IP, 并与地面核心网安全网关建立IPSec隧道的安全传输方式, 如图3所示。由于星载基站在持续高速移动, 需根据其运动轨迹来切换信关站, 此时星载基站需要重新获取公网IP并建立IPSec隧道。

信关站IPSec是指地面多个信关站通过公网与运营商核心网连接, 并建立多条IPSec隧道, 以此屏蔽信关站之间的差异, 如图4所示。星载基站分配运营商内网IP, 通过卫星平台提供的馈线链路通道连接信关站, 然后信关站通过IP-Sec隧道转发星载基站数据。同时, 馈线链路作为空间无线

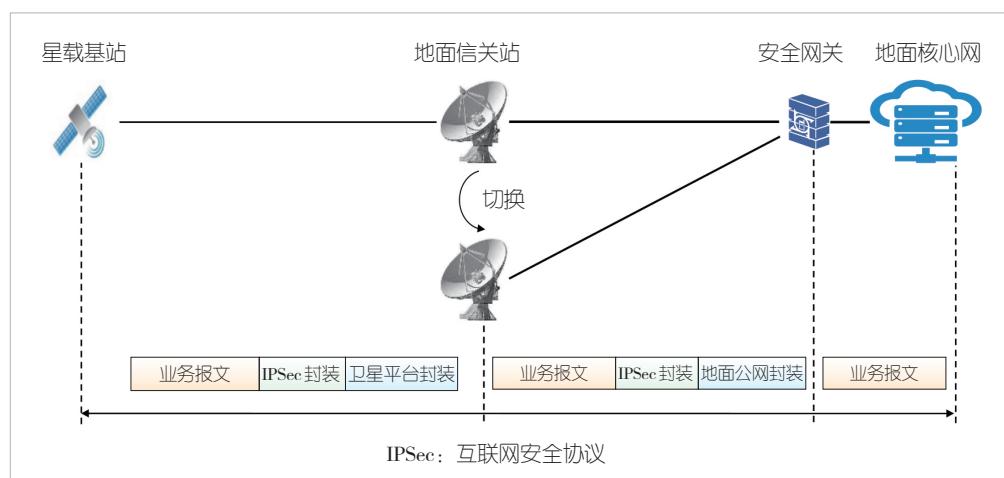


图3 基站IPSec星地回传安全传输

链路，增强了空口安全防护机制，以抵御空口干扰、窃听、伪造劫持等攻击。

3) 星间链路安全传输技术

天地一体化网络中的星间链路是星载基站、星载核心网、中高轨中继卫星之间的通信链路。它将多颗卫星互联，实现卫星间高速传输和交换，形成一个以卫星为交换节点的空间通信网络，并支持3GPP NG/Xn接口连接。星间链路的引入，使低轨卫星移动通信系统能够更少地依赖地面网络，更为灵活地进行路由选择和网络管理，从而有助于减少地面信关站数量，降低部署复杂度和成本。

激光通信技术应用于星间链路传输，相比微波通信有很大优势：首先是通信速率高，可不受无线频谱限制，传输速率可达到100 Gbit/s以上；同时还有抗干扰能力强、保密性好、设备轻量化及功耗低的优点。其缺点是异轨激光通信相对同轨通信，存在卫星相对移动速度快、对准难度高的问题。因此需要提供同轨与异轨场景下的高效连接管理机制，针对高轨核心网与低轨基站的NG接口连接，以及分布在低轨不同层次的星载基站之间的Xn口连接，支持根据网络拓扑变化实时调整激光对准精度。

此外，网元还可能部署在同一颗卫星或不同卫星上。例如，星载基站采用CU/DU分离部署、UPF下沉星载基站部署等。当这些网元位于不同卫星时，需为网元之间提供安全保障，包括网元间的信任关系建立和通信安全防护。

4) 轻量化无线物理层安全技术

在天地一体化网络中，由于庞大的星载基站处于持续高速移动状态，空口信道特征呈现高度动态变化，具有较强的随机性和时变性。这为以信道特征为基础的无线物理层安全(PLS)技术提供了应用优势。PLS技术以信道状态、信号强度、功率谱密度、信道冲激响应和方向角等信道属性为内生安全因子，形成信道指纹，是一种利用合法通信双方的信道的唯一性和互易性实现安全保护的内生安全机制；而传统安全机制则基于高层密钥协商和计算复杂度加解密算法实现。由于无线物理层安全机制大幅减少甚至没有高层信令交互，因此可降低信令处理时延和计算复杂度，支持更多用户终端和接入，并降低系统计算资源消耗，是比较理想的轻量化安全技术。此外，基于信道指纹的时变性可以实现接近“一次

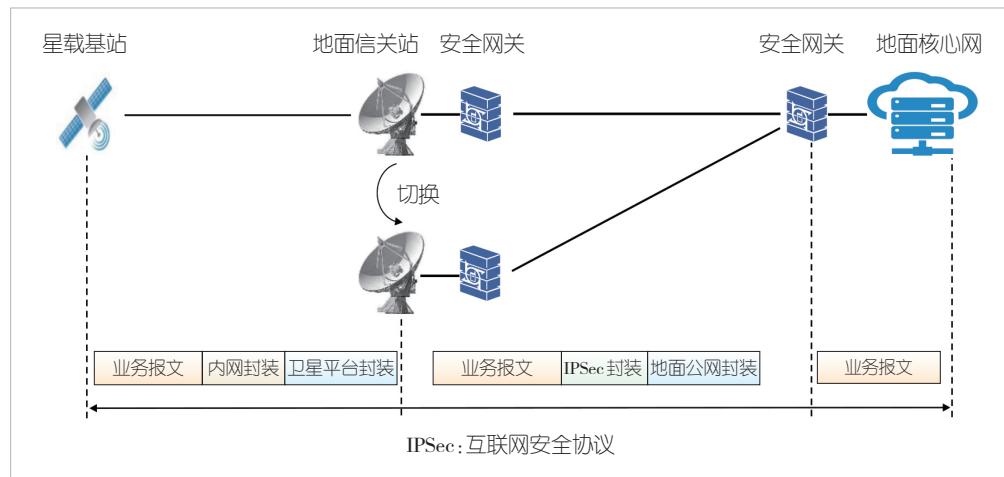


图4 信关站IPSec星地回传安全传输

一密”的安全效果，理论上是最安全的密码技术，有望成为抵抗量子攻击的候选关键技术。

除信道指纹外，无线物理层安全技术还可以辅助利用3类内生安全属性以进一步增强安全机制：一是基于发射机部件容差和工艺条件差异等形成的唯一射频指纹；二是基于天线制造公差产生的波束图形中不同的波束指纹；三是基于智能超表面(RIS)反射系数矩阵控制多径信号相移与幅度产生的相移指纹。

在天地一体化网络场景中，可以使用信道指纹和射频指纹进行轻量化无线物理层接入的身份检测和认证。网络侧利用信道互易性核对终端信道指纹后，即可判定终端的合法性，无需传统的交互认证协议。文献[13]提出了一种基于信道指纹的身份欺骗攻击检测方案，将波束域特性作为一种信道指纹，并将欺骗攻击中的身份检测问题建模成对其信道指纹的二分类问题，同时选用基于监督学习的支持向量机(SVM)算法。这种方案具有较高的认证准确率，对于不同频率、天线数量、移动速度(低速)的具有较好的鲁棒性。仿真结果表明，即使终端移动速度仅为5 m/s，所提方案仍可以保持95%的认证准确率^[13]。

无线物理层安全技术也可以对抗空口信号窃听，实现无密钥安全防护。物理层安全预编码利用信道特征设计预编码矩阵，包括波束成形、人工噪声以及天线子集选择等。波束成形通过设计预编码矩阵调节发射天线阵列或其子阵列，将发送信号的能量集中到合法用户方向，以提高合法用户的信道条件，从而增强接收信号质量。人工噪声利用无线信道以及噪声内在的随机性，使得合法用户的信噪比高于窃听者信道，以保证合法用户的信噪比高于窃听者，从而达到安全传输的目的。天线子集选择在保证合法用户正常接收信号的同时，扰乱窃听者星座图，使窃听者无法准确解调信号^[14]。

5) 抗量子密码及量子保密通信技术

随着量子技术的飞速发展,传统基于计算复杂度的密码算法及系统的安全性面临被量子计算破解的威胁。虽然对称密码可以通过增加密钥长度来抵抗量子攻击,但非对称密码如RSA(Rivest-Shamir-Adleman)、椭圆曲线密码(ECC)算法基于大数分解和离散对数的数学问题,正面临着量子计算的本质威胁。1994年SHOR等提出了分解大整数和求解离散对数的量子算法^[15],使得RSA、ECC密码算法可能被破解。未来,如果量子计算获得成功,则目前广泛使用的基于RSA、ECC密码的公钥体系可能不再安全,因此抗量子密码技术也是6G网络安全研究的关键技术。美国国家标准与技术研究所(NIST)在2023年8月发布了抗量子密码学的初始公开标准草案,包括基于有结构格的公钥加密/密钥封装算法Crystals-Kyber^[16]、基于有结构化格的公钥签名算法Crystals-Dilithium^[17]和基于哈希的公钥签名SPHINCS+^[18]。此外,在NIST抗量子密码标准化项目中,中国学者主导并参与提交了多个后量子密码算法。其中,由中科院信息工程研究所路献辉教授团队设计的基于格的加解密(LAC)算法进入了第二轮^[19]。

一直以来,中国在量子保密通信领域领跑世界。2023年8月工业和信息化部发布了《基于IPSec协议的量子保密通信应用设备技术规范》^[20]等3项量子保密通信相关的通信行业标准,对量子保密通信产品设计和安全测评提供了权威指导,推动了有关设备产品的安全应用。

本文提出一种应用于天地一体化网络的量子保密通信技术方案,如图5所示。该方案通过量子加解密密钥与IPSec协议融合,提供天地一体化网络加解密安全服务。在星载基站与地面核心网NG接口处IPSec加解密通道的基础上,引入量子密钥后,星载基站和地面安全网关分别连接量子密钥系统,获得定期稳定的量子密钥,并通过“量子卫星+量子站”链路的密钥分发系统与量子密钥保持同步。该方案同时可以对IPSec协议进行改造,在密钥协商第一阶段增加量子加密服务协商通知载荷,实现

量子加密服务的协商、密钥属性的交换和密钥获取结果的通知;在第二阶段改造使用量子密钥对原始会话密钥进行摘要或异或运算,生成最终的会话密钥。这种方案对IPSec协议改动较少,遵循标准规范,有利于商用推广,但缺点是量子密钥只作为密钥参数,不能直接使用量子密钥作为加密密钥。另一种方案是不启用IPSec密钥协商流程,而是直接使用量子密钥作为IPSec的加密密钥。该方式极大简化了IPSec协议流程,有利于实现轻量化的安全协议,提高系统性能和并发能力,但缺点是对IPSec协议改造很大,还未实现标准化,商用推广比较困难。

6) 星地可信联盟的区块链技术

天地一体化网络提供空、天、地、海全域泛在覆盖和海量用户连接,实现全球服务、应急服务和通导遥一体化服务,是中国“一带一路”的战略性基础设施。星地可信联盟网络满足全球卫星星座统一运营和各国地面运营商自治运营的需求,提供适配各国家政策法规和卫星/地面行业生态圈的安全可信与经济适用的泛在连接基础设施,由星载通信网的空间段和各国家地面通信网的地面段组成,如图6所示。

星地可信联盟是星载通信网和各国家地面通信网(基站+核心网)共同建立的联盟网络。这种网络通过星地馈线链路建立联盟连接,基于区块链分布式账本构建的共建式共识信任安全和交易框架,将空间星载通信网和各国家地面网共同纳入联盟安全可信体系,满足多方身份互信、各种资源

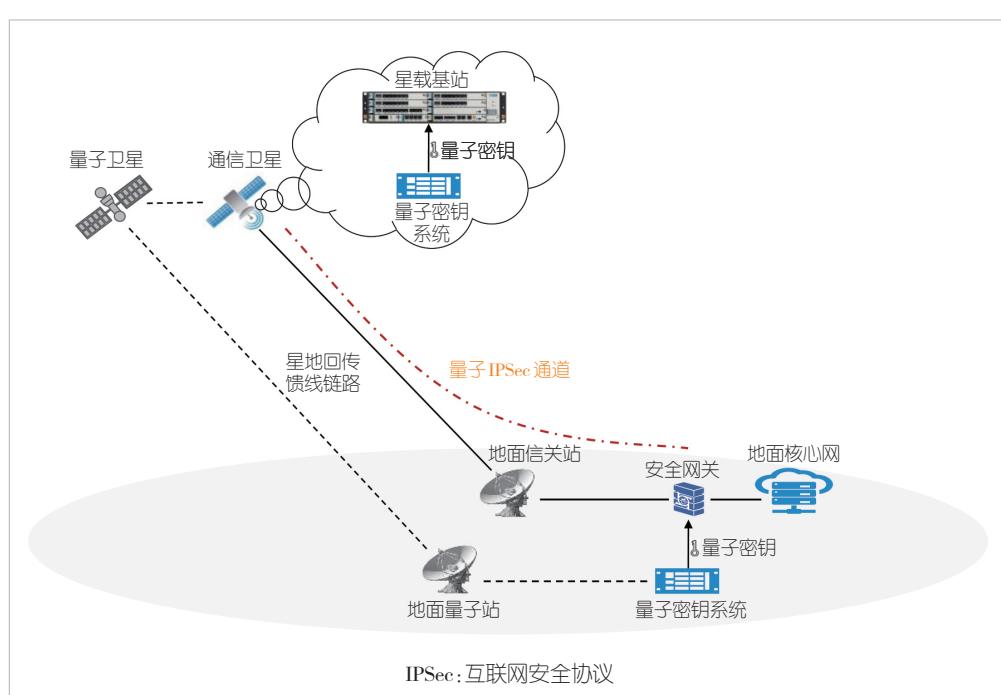


图5 天地一体化网络的量子保密通信技术方案

上链服务和价值兑现的需求，激活用户、行业和区域的参与度，打造商业模式创新平台。

区块链技术的一个突出特点是去中心化，即没有单一的实体可以控制联盟网络，联盟链上的每个节点都可以自主地维护自身的数据，能够有效避免中心化的安全漏洞。同时，所有传输数据都被加密并被添加到对应的区块中。每个区块都包含前一个区块的哈希值，形成了一个链式结构，其中每个节点都可以查看和验证数

据。由于每个区块都依赖前一个区块的数据，因此要想篡改其中一个区块，就需要同时修改所有后续区块，这样可有效防止数据篡改和欺诈行为。区块链技术使用加密算法来保护数据的安全，只有授权用户才能解密和查看数据，防止数据窃取和篡改。

区块链的智能合约用于执行预定义的规则和条件，并在满足特定条件时自动执行，可以有效减少人为因素对网络的影响，提高网络的安全性和可靠性。例如，当有新的卫星节点请求加入联盟网络时，智能合约会发起共识请求对卫星节点身份进行认证，联盟网络中所有节点达成共识后，会更新数据层区块链的链上数据，最后才允许新节点加入联盟网络。当地面站请求与联盟网络建立连接时，首先需要对区块链中的数据进行检索，以查看是否有该地面站的认证信息。如果已经通过认证，则允许建立连接；否则，需要经过节点共识才允许接入。

4 结束语

相比于5G，6G网络需要应对超高速、超低时延、超大规模连接场景带来的新的安全挑战。天地一体网络作为6G的基本组网架构，同样将面临新的安全风险，传统的网络安全技术也将难以适配。天地一体化网络面临的网络拓扑动态变化和卫星平台资源受限的安全挑战，将会极大影响6G网络研究和标准进程。因此，轻量化的接入认证、协议简化、轻量化的加解密和星地可信联盟等新技术研究是十分必要的。本文分析了接入认证、星地回传安全传输、星间链路安全传输、轻量化的无线物理层安全、抗量子密码及量子保密通信和基于区块链的星地可信联盟等网络安全技术，希望为后续6G天地一体化网络安全技术研究提供参考。

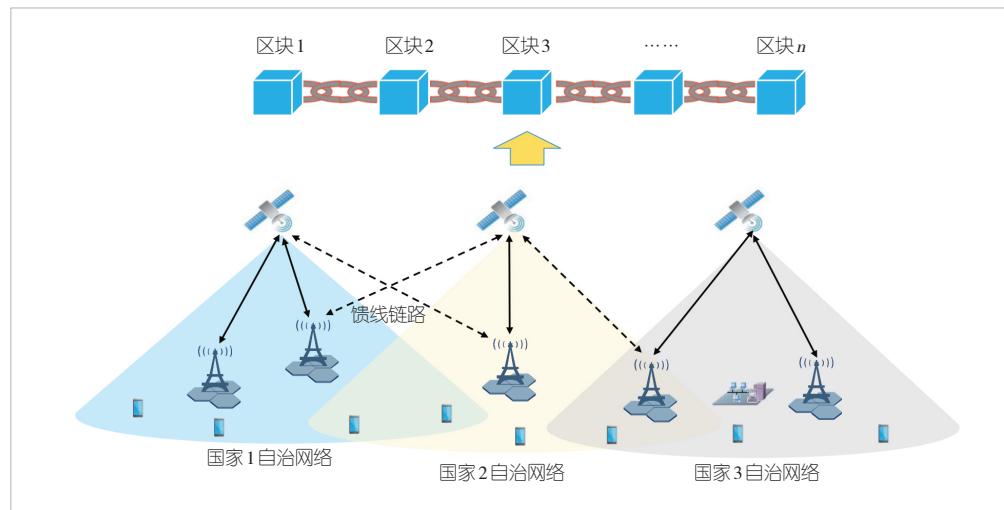


图6 天地一体化星地可信联盟

参考文献

- [1] 刁兆坤, 杨丽, 王振章. 6G空天地一体化网络架构及其构建 [J]. 通信世界, 2024(4): 36-39
- [2] 陈新宇, 张强, 陆光辉. 天地一体网络场景下的数字孪生关键技术 [J]. 中兴通讯技术, 2023, 27(2): 51-58. DOI: 10.12142/ZTETJ.202303010
- [3] 缪德山, 邓凌越, 孙建成, 等. 6G星地融合无线网络及关键技术 [J]. 中兴通讯技术, 2024, 30(4): 42-49. DOI: 10.12142/ZTETJ.202404007
- [4] 杨帅斌, 张昱, 卢为党. 面向6G的卫星通信感知一体化网络及关键技术 [J]. 中兴通讯技术, 2024, 30(5): 16-23. DOI: 10.12142/ZTETJ.202405004
- [5] 夏旭. 面向5G/6G卫星: NTN标准发展、关键技术与未来思考 [J]. 广播电视网络, 2024(5): 60-65
- [6] IMT-2030(6G)推进组. 6G网络安全愿景技术研究报告 [R]. 2021
- [7] 陆海涛, 陈一喆, 娄笃仕. 5G/5G-Advanced/6G接入网安全技术演进及内生安全 [J]. 中兴通讯技术, 2022, 28(6): 85-94. DOI: 10.12142/ZTETJ.202206014
- [8] 中国联通研究院. 6G网络安全需求及架构白皮书 [R]. 2024
- [9] 3GPP. Security architecture and procedures for 5G system (Release 15): 3GPP TS 33.501 [S]. 2019
- [10] LU H T, YAN X C, ZHOU Q, et al. Key intrinsic security technologies in 6G networks [J]. ZTE communications, 2022, 20 (4): 22-31. DOI: 10.12142/ZTECOM.202204004
- [11] 中国移动通信研究院. 构建可信数字身份体系筑牢数字文明基石白皮书 [R]. 2024
- [12] 徐晖, 陈山枝, 艾明. 面向6G的星地融合网络架构 [J]. 中兴通讯技术, 2023, 27(2): 9-15. DOI: 10.12142/ZTETJ.202305003
- [13] 杨立君, 李明航. 基于信道指纹的毫米波MIMO系统身份欺骗攻击检测方案 [J]. 电子与信息学报, 2023(12): 4228-4234
- [14] 倪云云, 陈伯庆, 李刚. 5G毫米波通信中的物理层安全预编码 [J]. 中兴通讯技术, 2021(4): 53-59. DOI: 10.12142/ZTETJ.202104011
- [15] SHOR P W. Algorithms for quantum computation: discrete logarithms and factoring [C]//Proceedings 35th Annual Symposium on Foundations of Computer Science. IEEE, 1994: 124-134. DOI: 10.1109/SFCS.1994.365700
- [16] NIST. Module-lattice-based key-encapsulation mechanism standard: 203 (draft) [EB/OL]. [2024-10-15]. <https://nvlpubs.nist.gov/nistpubs/FIPS/NIST.FIPS.203.pdf>
- [17] NIST. Module-lattice-based digital signature standard [EB/OL]. (2024-08-13) [2024-10-15]. <https://nvlpubs.nist.gov/nistpubs/FIPS/NIST.FIPS.204.pdf>
- [18] NIST. Stateless hash-based digital signature standard [EB/OL].

(2024-08-13) [2024-10-15]. <https://nvlpubs.nist.gov/nistpubs/fips/nist.fips.205.pdf>

[19] 西安电子科技大学广州研究院. 后量子密码迁移白皮书 [R]. 2024

[20] 中国通信标准化协会. 基于IPSec协议的量子保密通信应用设备技术规范: YD/T 4303-2023 [S]. 2023

作者简介



陆海涛, 中兴通讯股份有限公司资深系统架构师, 正高级工程师, CISSP; 主要从事无线网络架构、无线产品安全、大规模天线、动态频谱共享等技术研究; 牵头承担 10 余项国家科技重大专项、“863”计划课题, 获广东省科技进步奖; 发表论文 8 篇, 申请发明专利 60 余项。



周强 (通信作者), 中兴通讯股份有限公司 5G ToB 研发副总工, 高级工程师; 主要从事 3G/4G/5G 等无线通信技术研究; 拥有丰富的无线系统产品设计和研发经验, 牵头和参与多项国家科技重大专项课题; 申请发明专利 15 项。



代九龙, 中兴通讯股份有限公司 RAN 产品安全规划资深专家; 主要从事无线网络协议栈、无线网络安全架构、无线产品安全、无线网络产品行业应用等技术研究; 申请专利 8 项。



卢帆, 中兴通讯股份有限公司 RAN 中心产品安全总监, CISSP; 主要从事 5G/6G 网络安全、天地一体化网络、量子通信等无线通信技术研究; 申请发明专利 12 项。



李锐, 中兴通讯股份有限公司无线软件架构高级工程师; 主要从事无线网络架构、网络安全、网络传输、软件架构等技术研究和开发; 申请发明专利 30 余项。

《中兴通讯技术》第31卷总目次

卷 · 期 · 页

卷 · 期 · 页

卷首特稿

对AI加持网络重构的思考 邬贺铨 31-1-01

热点专题

6G 立体覆盖技术

- 专题导读 李建东, 刘俊宇 31-1-04
通感赋能低空覆盖网络技术 胡文豪, 何振耀, 徐甜甜, 许威 31-1-06
面向6G立体覆盖扩展的即插即用接入机制 刘睿, 韩书君, 许晓东 31-1-13
面向6G动态场景的可移动去蜂窝网络覆盖性能分析 翟浩雅, 郑家康, 章嘉懿, 艾渤 31-1-20
面向网络化通信感知一体化的立体覆盖方法 杨文彬, 邱裕鹤, 苏郁 31-1-28
面向6G典型场景的高能效立体覆盖技术研究 张吉发, 黄香港, 赵楠 31-1-39

智算网络演进及关键技术

- 专题导读 段晓东, 李丹, 虞红芳 31-2-01
星脉网络: 面向GPU集群集合通信与集中式路由的协同优化 李宝嘉, 何春志, 夏寅贲, 何泽坤, 王晓亮 31-2-03
智算中心Scale-Out网络的演进及GSE的实践 程伟强, 李新双, 白艳, 吕勇 31-2-14
基于大语言模型蒙特卡洛树搜索的智算网络故障根因分析系统 罗子秋, 苗宇铠, 李丹 31-2-21
算力智联网关键技术研究 易昕昕, 张乃晗, 刘雅承, 韩梦瑶, 曹畅 31-2-31

- 面向多算力中心协同的广域智算网络仿真架构设计 边彦晖, 刘明远, 虞红芳 31-2-39
在网计算: 愿景与关键技术 刘杨, 孙浩南, 程伟强 31-2-47

6G网络安全

- 专题导读 刘建伟, 王景璟 31-3-01
新一代电信云网内生安全架构研究 袁超颖, 白景鹏, 袁淑美, 何国锋 31-3-03
6G星地融合网络安全需求及关键技术 梁亚从, 徐晖 31-3-09
基于MITRE ATT&CK框架的卫星系统网络安全量化评估模型 朱飞, 曹进, 韩昌隆, 易威, 李晖 31-3-14
6G信任基础: 统一自我主权身份系统 李灏洋, 肖少荷, 曹宾, 彭木根, 李丽香, 刘欣然 31-3-22
AI赋能6G网络安全: 架构与关键技术 王瀚洲, 金子安, 王瑞, 刘建伟 31-3-31
人工智能赋能网络安全的挑战与应用 杨红梅, 赵勋 31-3-39
6G网络内生安全体系运行机制研究 粟粟, 闫茹, 马宇威 31-3-44
6G网络安全的架构与未来 罗涵一, 崔宝江, 全鑫 31-3-50

面向6G的高时效智能机器通信技术

- 专题导读 张平, 秦晓琦, 马楠 31-4-01
意图驱动的智简通信 戴金晟, 秦晓琦, 秦海龙, 张平 31-4-03
面向铁路入侵检测的语义通信技术

.....	郭疆远, 陈为, 艾渤	31 - 4 - 13
6G数字孪生信道的三个使能技术: 多模态感知、环境知识和大模型	于力, 张建华, 蔡逸辰	31 - 4 - 19
端边协同的6G内生AI网络	王志勤, 周吉皓, 韩凯峰	31 - 4 - 29
多智能体协作感知的现状与展望	刘根嘉, 陈思衡, 张文军	31 - 4 - 34
基于生成式模型的新型机器语义通信方法研究	承楠, 王秀程, 沈学民	31 - 4 - 41

网络中的AI技术

专题导读	解冲锋, 孟洛明, 崔勇	31 - 5 - 01
大模型驱动的网络智能运营管理标准化和应用展望	李文璟, 方宏林, 喻鹏	31 - 5 - 03
大语言模型赋能智能网络的应用与挑战	牛嘉林, 邢铭哲, 张蕾	31 - 5 - 11
AI智能体赋能网络运营的研究与应用	郑雨婷, 程新洲, 王静云	31 - 5 - 19
检索增强的网络流量预测方法	常远, 吴春鹏, 王峰	31 - 5 - 25
原生AI融合网络数字孪生赋能下一代无线网络自治	王首峰, 郭建超, 边森	31 - 5 - 30

新一代光传输技术

专题导读	陈建平, 唐雄燕	31 - 6 - 01
基于低复杂度Transformer的光纤信道快速精确建模技术	史明辉, 郑智雄, 牛泽坤, 义理林	31 - 6 - 03
面向智算中心互联的光算协同技术研究	谭艳霞, 满祥锟, 吴绍辉, 张贺, 徐博华	31 - 6 - 13
50G-PON平滑演进的挑战和方案研究	杨波, 张德智, 刘波	31 - 6 - 20
数字副载波复用光纤通感融合系统干涉衰落抑制研究	张兵兵, 向梦, 张泽彬, 付松年, 秦玉文	31 - 6 - 29
新一代光传输网络关键器件和集成化技术研究		

名家视点

面向低空经济的无线覆盖扩展技术思考与展望	盛敏, 陈旭晖, 马兴业, 赵晨曦	31 - 1 - 48
智算网络发展综述	段晓东, 程伟强, 张昊	31 - 2 - 53
6G网络安全架构展望	马红兵, 姚戈, 张曼君, 徐雷	31 - 3 - 56
6G: 从通信到通感算智融合	刘光毅	31 - 4 - 48
面向下一代光网络的光计算技术应用思考	李俊杰, 刘宇旸, 霍晓莉	31 - 4 - 55
光纤通信技术演进与发展展望: 从基础突破到融合创新	张海懿	31 - 5 - 37
关于AI原生的几点探讨	何宝宏	31 - 6 - 49

企业视界

通感低空覆盖关键技术与组网	林旭, 刘涛, 张诗壮, 施嵘	31 - 1 - 53
智算中心组网技术及应用	段威, 于浩, 李和松, 张征, 王宇诚	31 - 2 - 63
相变浸没式液冷系统研究	万积清, 柯文, 徐世杰, 朱冰	31 - 3 - 62
人工智能技术与应用前沿	包义明, 林阳, 屠要峰	31 - 4 - 64
下一代AI大模型计算范式洞察	熊先奎, 王程晨, 蔡文豪	31 - 5 - 50
智算光互联技术进展及趋势	张平化, 王会涛, 王勇	31 - 6 - 53

技术广角

基于X86通用硬件平台的高性能5G核心网UPF实现	李勇, 马建伟, 应晓冬, 娄方亮, 蔡茹鋆	31 - 1 - 58
---------------------------------	------------------------	-------------

智能实时音视频网络的总体设计	陈俊江, 申光	31-2-63
面向5G网络的工业互联网安全技术分析	陈焱, 陈丹, 袁琦, 刘小丽, 徐晓娜	31-2-72
物理层安全保密电话的实现	林立峰, 周子健, 焦秉立	31-2-77
云化IP城域网vBRAS可靠性研究	林在峰, 张文强, 杨冰, 龚霞, 朱永庆	31-3-68
星地一体化语义通信网络:探索与展望		

..... 李东博, 王新宇, 尹志胜, 承楠, 刘勘	31-5-57	
5G基站节能面临的关键问题和解决方案	王小锋, 韩茜	31-5-66
基于熵函数的语义信息边界研究	唐雪, 许进, 冯雨龙	31-6-61
智能超表面自动化部署解决方案	李萍, 叶峥峥, 窦建武	31-6-70
面向6G的天地一体化网络安全技术	陆海涛, 周强, 代九龙, 卢帆, 李锐	31-6-75

综合信息

《中兴通讯技术》2026年专题计划

期次	专题名称	策划人
1	6G关键技术的标准化:Day-1与未来	易芝玲 中国移动研究院首席科学家
2	广域立体覆盖低空通信技术	金石 东南大学副校长 刘凡 东南大学教授
3	智算网络	赵慧玲 工信部信息通信科技委常委
4	大模型推理中的存算技术	郑纬民 中国工程院院士、清华大学教授 陆游游 清华大学副教授
5	星地太赫兹高速传输技术	洪伟 中国科学院院士、东南大学教授 唐万斌 电子科技大学教授 郝张成 东南大学教授
6	智能多天线技术	艾渤 北京交通大学副校长 章嘉懿 北京交通大学教授

中兴通讯技术杂志社

促进产学研合作青年专家委员会

主任 陈 为(北京交通大学)

副主任 秦晓琦(北京邮电大学) 卢 丹(中兴通讯股份有限公司)

委员

曹 进 西安电子科技大学

陈 力 中国科学技术大学

陈 为 北京交通大学

陈琪美 武汉大学

陈舒怡 哈尔滨工业大学

陈思衡 上海交通大学

官 科 北京交通大学

韩凯峰 中国信息通信研究院

何 姿 南京理工大学

侯天为 北京交通大学

胡 杰 电子科技大学

黄 晨 紫金山实验室

李 昂 西安交通大学

刘 凡 东南大学

刘春森 复旦大学

刘俊宇 西安电子科技大学

卢 丹 中兴通讯股份有限公司

陆游游 清华大学

宁兆龙 重庆邮电大学

祁 亮 上海交通大学

秦晓琦 北京邮电大学

秦志金 清华大学

史颖欢 南京大学

唐万恺 东南大学

王景璟 北京航空航天大学

王兴刚 华中科技大学

王勇强 天津大学

温森文 华南理工大学

吴泳澎 上海交通大学

武庆庆 上海交通大学

夏文超 南京邮电大学

徐梦炜 北京邮电大学

徐天衡 中国科学院上海高等研究院

杨川川 北京大学

尹海帆 华中科技大学

于季弘 北京理工大学

张 娇 北京邮电大学

张宇超 北京邮电大学

章嘉懿 北京交通大学

赵昱达 浙江大学

赵中原 北京邮电大学

周 伊 西南交通大学

朱秉诚 东南大学

刊物相关信息



投稿须知



投稿平台



过刊下载



论文索引与
引用指南

办刊宗旨：

以人为本，荟萃通信技术领域精英
迎接挑战，把握世界通信技术动态
立即行动，求解通信发展疑难课题
励精图治，促进民族信息产业崛起

产业顾问：

段向阳、高 音、胡留军、华新海、刘新阳、
史伟强、屠要峰、王会涛、熊先奎、许 进、
闫新成、赵亚军、朱晓光

双月刊 1995 年创刊

第 31 卷 总第 186 期

2025 年 12 月 第 6 期 (卷终)

主管：安徽出版集团有限责任公司

主办：时代出版传媒股份有限公司

深圳航天广宇工业有限公司

出版：安徽科学技术出版社

编辑、发行：中兴通讯技术杂志社

总编辑：王喜瑜

主编：陶善勇

执行主编：黄新明

副主编：卢丹

编辑部主任：王萍萍

责任编辑：徐烨

编辑：杨广西、朱莉、任溪溪

设计排版：徐莹

发行：王萍萍

编务：王坤

《中兴通讯技术》编辑部

地址：合肥市金寨路 329 号凯旋大厦 1201 室

邮编：230061

网址：tech.zte.com.cn

投稿平台：tech.zte.com.cn/submission

电子邮箱：magazine@zte.com.cn

电话：(0551) 65533356

发行方式：自办发行

印刷：安徽添锦印刷科技有限公司

出版日期：2025 年 12 月 30 日

中国标准连续出版物号：ISSN 1009-6868
CN 34-1228/TN

定价：每册 20.00 元