



信息通信领域产学研合作特色期刊 十佳皖刊
第三届国家期刊奖百种重点期刊 中国科技核心期刊

ISSN 1009-6868
CN 34-1228/TN

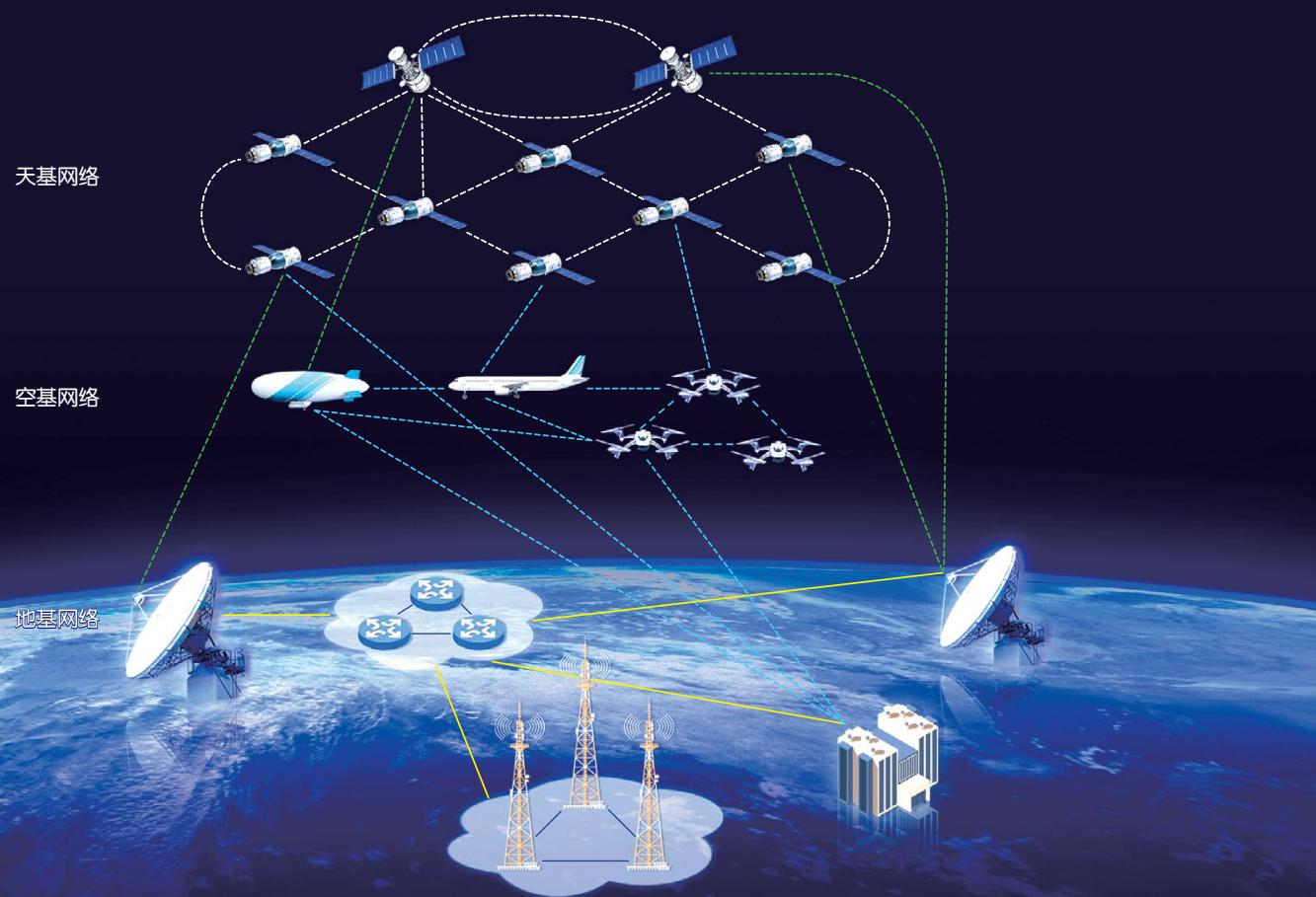
中兴通讯技术

ZTE TECHNOLOGY JOURNAL

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

2023年10月·第5期

专题: 6G 网络技术



ISSN 1009-6868



9 771009 686236

(封面图片详解见 P25)

《中兴通讯技术》第9届编辑委员会成员名单

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)

陈锡生(南京邮电大学教授) 麋正琨(南京邮电大学教授)

主任 陆建华(中国科学院院士)

副主任 李自学(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 (按姓名拼音排序)

陈建平 上海交通大学教授

陈前斌 重庆邮电大学教授、副校长

段晓东 中国移动研究院副院长

葛建华 西安电子科技大学教授

管海兵 上海交通大学教授

郭 庆 哈尔滨工业大学教授

洪 伟 东南大学教授

黄宇红 中国移动研究院院长

纪越峰 北京邮电大学教授

江 涛 华中科技大学教授

蒋林涛 中国信息通信研究院科技委主任

金 石 东南大学首席教授、副校长

李尔平 浙江大学教授

李红滨 北京大学教授

李厚强 中国科学技术大学教授

李建东 西安电子科技大学教授

李乐民 中国工程院院士

李融林 华南理工大学教授

李自学 中兴通讯股份有限公司董事长

林晓东 中兴通讯股份有限公司副总裁

刘 健 中兴通讯股份有限公司高级副总裁

刘建伟 北京航空航天大学教授

隆克平 北京科技大学教授

陆建华 中国科学院院士

马建国 之江实验室教授

毛军发 中国科学院院士

孟洛明 北京邮电大学教授

石光明 鹏城实验室副主任

孙知信 南京邮电大学教授

谈振辉 北京交通大学教授

唐 宏 中国电信IP领域首席专家

唐雄燕 中国联通研究院副院长

陶小峰 北京邮电大学教授

王文博 北京邮电大学教授、副校长

王文东 北京邮电大学教授

王喜瑜 中兴通讯股份有限公司执行副总裁

王 翔 中兴通讯股份有限公司高级副总裁

王耀南 中国工程院院士

王志勤 中国信息通信研究院副院长

卫 国 中国科学技术大学教授

吴春明 浙江大学教授

邬贺铨 中国工程院院士

向际鹰 中兴通讯股份有限公司首席科学家

肖 甫 南京邮电大学教授、副校长

解冲锋 中国电信研究院教授级高工

徐安士 北京大学教授

徐子阳 中兴通讯股份有限公司总裁

续合元 中国信息通信研究院副总工

薛向阳 复旦大学教授

薛一波 清华大学教授

杨义先 北京邮电大学教授

叶 茂 电子科技大学教授

易芝玲 中国移动研究院首席科学家

张宏科 中国工程院院士

张 平 中国工程院院士

张钦宇 哈尔滨工业大学(深圳)教授、副校长

张 卫 复旦大学教授

张云勇 中国联通云南分公司总经理

赵慧玲 工业和信息化部信息通信科技委常委

郑纬民 中国工程院院士

钟章队 北京交通大学教授

周 亮 南京邮电大学教授、副校长

朱近康 中国科学技术大学教授

祝宁华 中国科学院院士

目 次

中兴通讯技术 (ZHONGXING TONGXUN JISHU)
总第 172 期 第 29 卷 第 5 期 2023 年 10 月

信息通信领域产学研合作特色期刊 第三届国家期刊奖百种重点期刊 中国科技核心期刊 工信部优秀科技期刊 十佳院刊 中国五大文献数据库收录期刊 1995 年创刊

热点专题 ▶

6G 网络技术

- 01 专题导读 王文东
02 6G 智能内生网络架构及关键技术分析 李文璟, 喻鹏, 张平
09 面向 6G 的星地融合网络架构 徐晖, 陈山枝, 艾明
16 6G 网络架构展望 刘玉芹, 邢燕霞, 陈鹏
21 6G 网络架构和关键技术展望 王友祥, 唐雄燕
28 6G 网络架构研究进展及建议 谢峰
38 面向 6G 无线环境可预测的感知重构、语义表征及应用 张建华, 田艺璇, 孙语瞳, 于力
45 面向下一代移动通信的前传网络关键技术 赵俊皓, 张俊文, 迟楠
56 基于机器学习的智能路由解释方法 孟子立, 徐明伟
61 网络智能传输研究进展 廖乙鑫, 王子逸, 崔勇
68 NetGPT:超越个性化生成服务的内生智能网络架构 陈宇轩, 李荣鹏, 张宏纲

专家论坛 ▶

- 76 6G 新型信息通信网络架构设计 段晓东, 孙滔, 刘超, 施南翔

企业视界 ▶

- 82 相干光收发器件未来技术演进 沈百林, 王会涛

综合信息 ▶

- 37 《中兴通讯技术》2024 年专题计划

《中兴通讯技术》2023 年热点专题名称及策划人

1. 面向云网安全的新型防护技术

中国电信研究院教授级高工 解冲锋
北京邮电大学教授 杨义先

3. 数字孪生技术

重庆邮电大学教授 陈前斌

5. 6G 网络技术

北京邮电大学教授 王文东

2. 语义通信

中国科学院院士 陆建华
清华大学教授 陶晓明

4. 算力网络和东数西算

工业和信息化部信息通信
科技委常委 赵慧玲

6. 面向双碳的新一代无线通信网络

华中科技大学教授 葛晓虎
西安电子科技大学教授 李建东

MAIN CONTENTS

ZTE TECHNOLOGY JOURNAL
Vol. 29 No. 5 Oct. 2023

Special Topic ►

6G Network Technology

- 01 Editorial WANG Wendong
- 02 Architecture and Key Technologies of 6G Intelligent Endogenous Network
..... LI Wenjing, YU Peng, ZHANG Ping
- 09 Integrated Satellite and Terrestrial Network Architecture for 6G
..... XU Hui, CHEN Shanzhi, AI Ming
- 16 Prospects for 6G Network Architecture LIU Yuqin, XING Yanxia, CHEN Peng
- 21 Prospect of 6G Network Architecture and Key Technologies
..... WANG Youxiang, TANG Xiongyan
- 28 6G Network Architecture: A Survey XIE Feng
- 38 Sensing Reconstruction, Semantic Representation and Application Towards 6G Predictive
Wireless Environment ZHANG Jianhua, TIAN Yixuan, SUN Yutong, YU Li
- 45 Key Technologies for Fronthaul Network in Next-Generation Mobile Communication
..... ZHAO Junhao, ZHANG Junwen, CHI Nan
- 56 Interpreting Machine Learning-Based Intelligent Routing Algorithms
..... MENG Zili, XU Mingwei
- 61 Network Intelligent Transmission Technology: A Survey
..... LIAO Yixin, WANG Ziyi, CUI Yong
- 68 NetGPT: An AI-Native Network Architecture for Provisioning Beyond Personalized Generative
Services CHEN Yuxuan, LI Rongpeng, ZHANG Honggang
- 76 Architecture Design of 6G New Information Communication Network
..... DUAN Xiaodong, SUN Tao, LIU Chao, SHI Nanxiang
- 82 Future Technology Evolution of Coherent Transceiver Optical Devices
..... SHEN Bailin, WANG Huitao

Expert Forum ►

Enterprise View ►

期刊基本参数：CN 34-1228/TN*1995*b*16*86*zh*p*¥20.00*6500*13*2023-10

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。

6G 网络技术专题导读



专题策划人



王文东，北京邮电大学计算机学院网络与交换技术国家重点实验室教授、博士生导师；长期从事计算机网络理论和技术方面的教学和科研工作；获得茅以升北京青年科技奖，入选教育部“新世纪优秀人才支持计划”，享受国务院“政府特殊津贴”；相关科研和教学成果获国家和省部级以上奖励8次；发表学术论文200余篇，合作发布IETF标准1项，拥有国际发明专利多项、中国发明专利70余项。

近年来随着6G研究热情的持续高涨，世界许多国家和地区均已启动6G的研究计划。中国也在“十四五”规划纲要中明确提出，要前瞻布局6G网络技术的研究，推进6G技术的各项研究工作。目前，关于6G网络的研究工作主要集中在通信与人工智能（AI）深度融合的智能内生网络、空天地一体化的星地融合通信网络、通用高效的无线/有线连接和极致性能体验的智简网络等方面。本期专题和专家论坛栏目以6G网络技术为主题，邀请该领域的专家学者撰写了11篇文章。这些文章对6G网络技术的愿景、主要技术挑战进行了介绍、分析，也对6G研究领域涉及的部分技术内容进行了阐述。

《6G智能内生网络架构及关键技术分析》综述了6G智能内生网络的概念和特征，对数据与知识双驱动的网络架构和以任务为中心的网络架构进行了分析，对全球标准化组织在该架构上的标准化现状和问题进行了总结，并结合6G网络面临的挑战，阐述了6G智能内生网络的关键技术；《面向6G的星地融合网络架构》回顾了星地融合网络的研究现状和6G星地融合网络的发展趋势，介绍了针对全域覆盖、移动性管理、组网等方面的6G星地融合网络需求，分析了6G网络在星地融合挑战下的技术发展和应用前景，阐述了智简赋能的6G网络体系架构和弹性可重构的6G星地融合架构及相关技术；《6G网络架构展望》《6G网络架构和关键技术展望》和《6G网络架构研究进展及建议》3篇文章从不同角度对6G网络架构相关的研究工作进展和关键技术进行了全面的分析和介绍，并对中国未来在6G网络方面的研究工作进

行了展望；《面向6G无线环境可预测的感知重构、语义表征及应用》总结了无线环境可预测的语义表征方法，围绕信道建模与预测的难题，介绍了信道在线预测的6G网络设计与感知重构平台，并对未来无线环境可预测的研究进行了展望；《面向下一代移动通信的前传网络关键技术》以光纤无线融合接入为基础，针对数字前传技术、模拟前传技术和数模结合前传技术等，围绕网络架构、关键技术和未来发展方向进行了分析和探讨，对6G前传网络关键技术进行了展望；《基于机器学习的智能路由解释方法》综述了基于机器学习的智能路由方法进展，提出了一种可以对神经网络等黑盒子技术的输出决策结果进行解释的智能路由技术解释方法；《网络智能传输研究进展》针对6G网络和人工智能技术的发展趋势，对机器学习技术与传输层、应用层流媒体传输相结合的智能传输协议进行了分析，从网络传输协议和人工智能结合等方面展望了网络智能传输面临的机遇与挑战。《Net-GPT：超越个性化生成服务的内生智能网络架构》提出了基于边缘和云端部署相匹配大型语言模型（LLM）的NetGPT方案，展示了NetGPT为智能网络管理和编排提供一个统一解决方案的能力。专家论坛栏目文章《6G新型信息通信网络架构设计》，阐述了6G网络架构设计的6个原则和5个维度，进一步提出了“三体四层五面”的6G总体架构，以及端到端的全服务化系统架构和灵活按需的分布式自治组网架构，讨论了6G与5G网络架构的关系，并给出了未来研究方向及相关产业发展建议。

本期的作者来自知名高校与科研机构。面向6G网络技术，文章从技术挑战、标准化进展、系统架构和关键技术等方面介绍了最新的研究成果。希望本期的内容能为读者提供有益的启示和参考，并在此对所有作者的大力支持和审稿专家的辛勤指导表示由衷的感谢！

6G 智能内生网络架构及关键技术分析



Architecture and Key Technologies of 6G Intelligent Endogenous Network

李文璟/LI Wenjing, 喻鹏/YU Peng, 张平/ZHANG Ping

(北京邮电大学网络与交换技术全国重点实验室, 中国 北京 100876)
(State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing 100876, China)

DOI: 10.12142/ZTETJ.202305002

网络出版地址: <http://kns.cnki.net/kcms/detail/34.1228.TN.20231017.0948.002.html>

网络出版日期: 2023-10-18

收稿日期: 2023-08-05

摘要: 智能内生已成为6G网络的重要特征之一,也是当前业界关注的重点。首先解析了6G智能内生网络(IEN)的概念和特征,并对分层分面设计理念和典型架构进行了分析。之后,总结和展望了6G IEN架构的标准化进展。结合这些研究和进展,总结6G IEN面临的挑战,从知识表征与构建、意图驱动、分布式人工智能(AI)和AI可解释性4个角度分析了6G IEN的关键技术,为6G IEN架构的进一步演进提供了参考依据。

关键词: 6G; IEN; AI; 网络架构

Abstract: The intelligent endogenous network has become one of the important characteristics of the 6G network, and is also the focus of the current industry. The concept and characteristics of 6G intelligent endogenous network (IEN) are analyzed, as well as the layered design concepts and typical architectures of 6G IEN. Then the standardization progress of 6G IEN architecture is summarized. Based on research and standardization progress, the challenges faced by 6G IEN are summarized, and the key technologies are analyzed from the perspectives of knowledge characterization and construction, intention-driven, distributed artificial intelligence (AI), and AI interpretability, which provides a reference for the further evolution of 6G intelligent endogenous network architecture.

Keywords: 6G; IEN; AI; network architecture

引用格式: 李文璟, 喻鹏, 张平. 6G智能内生网络架构及关键技术分析 [J]. 中兴通讯技术, 2023, 29(5): 2–8. DOI: 10.12142/ZTETJ.202305002

Citation: LI W J, YU P, ZHANG P. Architecture and key technologies of 6G intelligent endogenous network [J]. ZTE technology journal, 2023, 29(5): 2–8. DOI: 10.12142/ZTETJ.202305002

未 来,人类将进入智能化时代。6G将提供智能化时代的网络连接服务,实现人机物间的智能互联和智慧协同。与人工智能(AI)技术深度融合,构建智能内生网络(IEN)已成为6G网络的重要特征。

目前5G网络的很多方面已经应用了AI技术。但是5G网络在架构设计之初,并未考虑AI技术,因此其智能化应用是在传统网络架构上进行优化和改造的,总体属于“外挂式智能”。外挂式智能是指采用AI技术在某个或某些方面提高网络运行的智能化水平,而不是从架构层面根本性地引入智能化。这种外挂式智能只是对传统架构和技术的局部增强,但在全面提升网络的智能化及自治水平方面就显得力不

从心。因此6G网络需要在架构设计之初就全面考虑与智能化的深度融合,构建架构级智能内生,使智能成为网络的“基因”,从而实现以AI为基础的新型智能通信网络。

本文首先提出了6G智能内生网络的概念和特征,然后对6G智能内生网络架构在产业界、学术界的研究和标准化现状、6G智能内生网络的关键技术进行了分析,为未来6G智能化网络架构的演进提供了一定的参考依据。

1 智能内生网络概念及特征

目前,学术界和产业界对智能内生概念并没有统一的定义。相对于外挂式智能,6G智能内生网络是与智能要素深度融合的网络,通过对智能要素的分布式编排、部署和协同调度,按需实现对内对外的智能服务。6G智能内生网络通过在端到端系统引入智能能力,助力网络实现无人工干预的

基金项目: 国家自然科学基金联合基金重点项目(U22B2031)

自感知、自分析、自决策、自执行和自评估的闭环自治。

6G智能内生网络将体现出分层分布式智能、智能要素服务化、服务自适应构建与开放性，以及开放安全性等特征。

1) 分层分布式智能：为满足未来6G全息化、沉浸式、虚实结合的业务需求，解决算力、时延、传输能耗、数据隐私等问题，智能要素的下沉部署已成为趋势，智能化逐渐从集中式向分布式发展。同时，集中式管控与分布式协作形成分层智能，共同实现跨域、跨网络的智能协同，满足不同智能业务的差异化需求。

2) 智能要素服务化：通过对智能要素（算力、算法和数据）的模块化设计，实现智能要素的独立扩展演进和灵活编排部署，并可通过服务的形式独立或联合对外提供智能化能力，达到智能化能力的高效复用。

3) 服务自适应构建与开放性：智能内生网络能够自主感知业务及网络运营需求，针对不同行业、用户及场景差异，自主实现对智能要素的设计、编排和调度，按需实现服务的自适应构建，精准匹配业务需求；同时，还可以支持服务的对外开放，向内外服务消费者提供各类服务。

4) 开放安全性：未来6G服务场景的多样化和泛在化，使得智能内生网络提供的服务能被各类服务消费者安全可信地订购和调用，并能保证智能服务的隐私性、鲁棒性和可解释性。

2 智能内生网络架构研究现状

面向6G网络智能化发展需求，为支持6G智能内生网络的特征和实现目标，产业界和学术界对智能内生网络架构进行了广泛的研究和探讨，同时6G智能内生网络架构的标准化工作也在持续推进中。

2.1 产业界与学术界的研究现状

为了支持6G网络的智能内生，当前产业界和学术界从网络架构角度，提出了新的智能平面和网元功能。

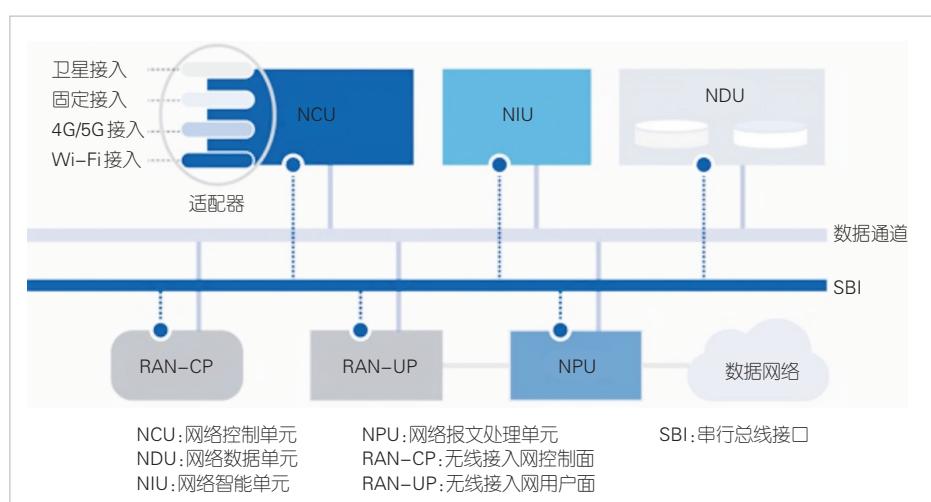
中国移动提出了“三体四层五面”的6G总体架构，从空间、逻辑与功能组成3个角度呈现了跨域、跨层、多维的6G网络。“三体”为网络本体、管理编排体、数字孪生体，“四层”为资源与算力层、路由与连接层、服务化功能层、开放使能层，

“五面”为控制面、用户面、数据面、智能面、安全面。其中，智能面是实现智能服务的物理载体，既可以为服务对象提供本地AI能力，又可以通过分布式智能节点的协同提供全局AI能力。安全面以“安全数据+AI”为驱动的安全感知和主动防护，构筑零信任的安全体系，实现安全内生^[1]。

中国电信和中兴通讯联合提出“三层四面”的6G网络智能化架构，包括云网资源层、网络功能层和应用使能层，以及控制面、用户面、数据面和智能面。与“四面”相对应，网络架构简化成四大网络单元：网络控制单元（NCU）、网络报文处理单元（NPU）、网络数据单元（NDU）和网络智能单元（NIU），各网络单元的互联采用串行总线接口（SBI）和数据通道双总线，改进了现有网络网元数量多、信令复杂等问题，实现灵活可定制的网络。SBI是现有的服务化信令交互总线，延续了已有服务化设计理念，将服务化范围由核心网控制面发展至用户面，同时进一步向无线接入网（RAN）延伸，包括RAN控制面（RAN-CP）和RAN用户面（RAN-UP），实现移动领域端到端的服务化；数据通道为新增数据总线，不同网络单元通过数据通道与NDU交互，以达到数据的高效获取与处理^[2]，具体如图1所示。

该架构呈现智能内生、分布式协同的特点，即“独立AI+内嵌AI”混合模式。NCU、NPU、NDU等网络单元一方面通过内嵌NIU服务，具备数据处理和AI分析能力，同时与独立NIU组成立体式学习和推理框架，按需、动态调配各NIU的数据、能力和算力资源，共同构建多点协同的智能内生网络^[2]。

从智能化能力的生成和演进方式来看，当前以规则式算法为核心的网络运行机理受限于通信-感知-计算能力和刚性预设式的规则，很难动态适配持续变化的用户需求和网络



▲图1 数据驱动的分布自治架构

环境。为了使网络有效累积并自主利用运行经验，实现网络运行及管控能力的自主迭代演进，文献[3]提出了以知识为中心的智能内生网络架构，如图2所示^[3]。

该架构内层是自进化核，或称为“知识大脑”，即以知识为核心，实现知识的获取、分析和更新，将网络的运行和维护经验以知识的方式识别并积累，从而为网络的自进化提供智能基础。外层是面向内生智能的网络闭环运行与管理功能，以实现网络的自主运行和自主维护，包含感知、规划、部署、运行、评估5个阶段，通过感知向自进化核提供数据和经验输入，根据自进化核提供的知识实现网络的规划、部署、运行和评估。

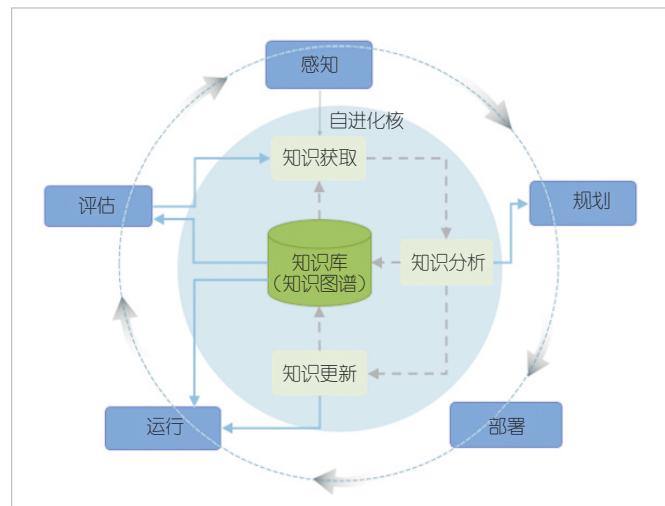
为了改善数据驱动的深度模型数据获取成本高、可解释性弱、鲁棒性不强等问题，文献[4]提出引入逻辑、视觉、物理定律和因果等知识，建立以海量数据为驱动、以知识为引导和约束的AI方法，为智能内生网络提供丰富的知识库。

为了提升智能化程度，文献[5]引入了意图抽象平面和认知平面，提出了意图抽象与知识联合驱动的6G内生智能网络架构，使6G网络支持感知-通信-决策-控制能力，能够自主感知周围环境及应用服务特性，从而进行自动化决策与闭环控制。该架构的目标是实现网络零接触、可交互、会学习，包括意图抽象平面、认知平面、管理平面、控制平面和数据平面，具体如图3所示^[5]。

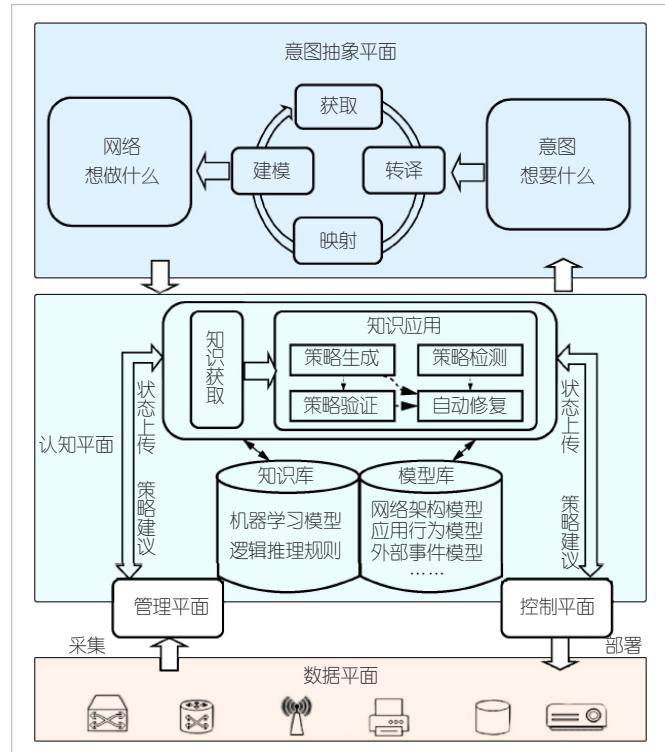
意图抽象平面的目标是从“想要什么”转换为“想做什么”；认知平面的目标是从“想做什么”转换为“怎么配置”，这需要通过知识获取和知识应用来实现。因此，本文所提的架构融合了基于意图的网络与知识定义网络的优势：首先抽象用户意图和感知网络状态，然后基于网络知识进行管控策略的生成与部署，最终实现内生智能网络的目标。

从智能要素服务化角度，华为等多家单位联合提出了一种以任务为中心的6G网络AI架构，引入了AI管理单元来管理AI任务，如图4所示。为支持网络原生AI能力，网络架构角度有3个变化：无线网络系统管控对象从“会话”变为“任务”；调度资源从连接资源变为连接、计算、数据和算法四要素资源，实现智能要素服务化；基于任务粒度的管控需要实现四要素的深度协同^[6]。

该架构包括网络AI管理编排（NAMO）和任务管控两部分核心内容。其中，NAMO完成从AI业务到任务的分解、映射和AI业务流编排，通常非实时，一般部署于管理域；而任务管控则完成任务实例的部署、控制、执行及生命周期管理，一般部署在控制面，以确保任务控制的实时和高效。该架构通过在无线通信系统中原生集成和融合四要素协同能力，并通过NAMO和任务管控的协同，最终保障了网络AI



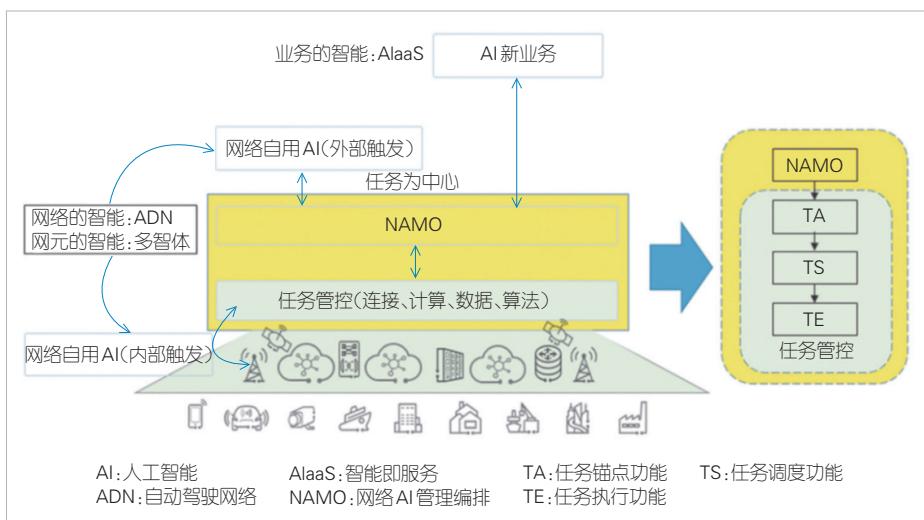
▲图2 双层闭环的网络运行与管控架构图^[3]



▲图3 意图抽象与知识联合驱动的6G内生智能网络架构^[5]

服务质量（QoAIS）^[6]。

综上所述，目前对6G智能内生网络架构的广泛探讨从分层分面的总体架构、智能能力生成、智能能力提升、AI能力编排与管理、分布式AI单元协作等方面展开，这体现了6G智能内生网络支持分层分布式智能、智能要素服务化、服务自适应构建，以及开放安全的特征。但当前关于6G智能内生网络架构的研究仍处于探索阶段。研究人员对6G智能内生网络架构进行自由探索，主要侧重于功能设计、架构

▲图4 以任务为中心的逻辑架构^[6]

设计,以及小规模、部分功能的原型系统实现,尚缺少大规模、全功能的实现支撑。

2.2 标准化研究现状

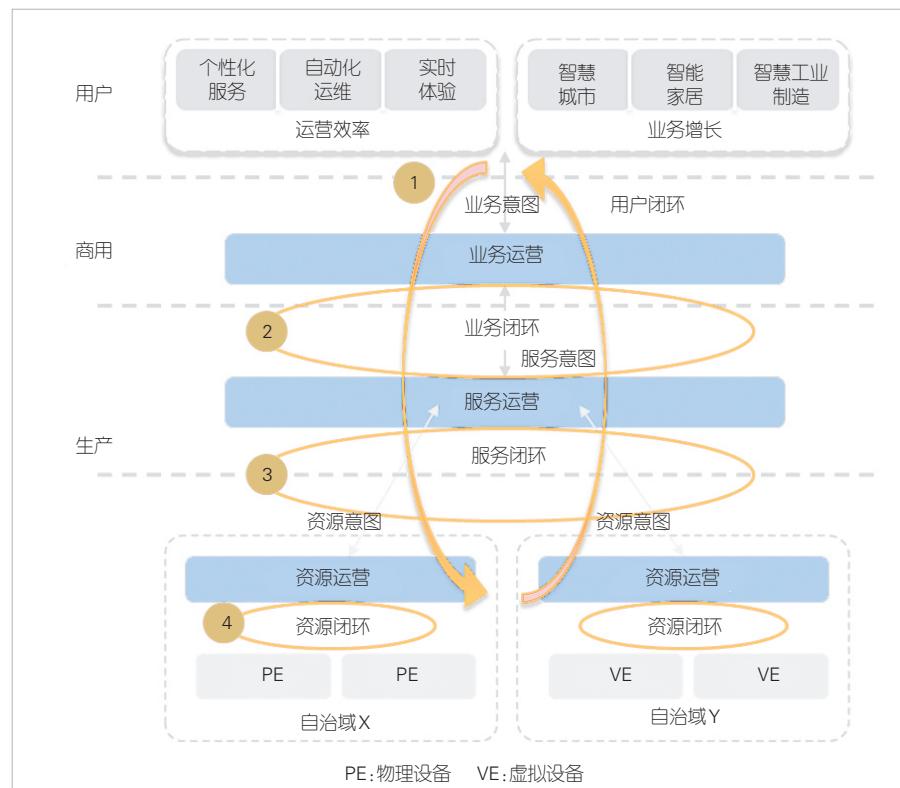
面向6G通信场景复杂化、业务需求多样化、业务体验个性化的需求,6G网络如何与AI相结合在不断地探索和演进中。欧洲电信标准化协会(ETSI)、电信管理论坛(TM Forum)、第3代合作伙伴计划(3GPP)、国际电信联盟电信标准分局(ITU-T)和IMT-2030(6G)推进组等标准化组织或论坛均取得了一系列进展。

ETSI很早就开始了网络与AI结合领域的探索,2017年就成立了体验式网络智能(ENI)工作组,致力于构建基于数据驱动决策和闭环控制的人工智能网络体系架构。ENI定义了一种用于网络运维、业务编排、网络保障等应用,提供基于模型、策略驱动、上下文感知的智能化服务的人工智能引擎,其功能架构包含知识管理、模型管理、策略管理等模块,通过对数据进行处理,利用AI模型可以为网络提供自动化服务运营和保障,并提供切片管理和资源编排。ENI系统基于体验式架构,通过自学习原则积累经验以持续提高运营效率,并使系统能够随着时间的推移实现提议到实施决策的全流程闭环控制。目前,ENI的功能还在不断地演进和丰富,如

支持基于意图驱动的网络等^[7]。

TM Forum提出了自治网络框架,该框架分为3个层级和4个闭环,如图5所示。其中,3个层级为通用运营能力,可支撑所有场景和业务需求,其中包括:资源运营层,面向单个自治域提供网络资源和能力自动化;服务运营层,面向多个自治域提供信息通信(IT)服务、网络规划、设计、上线、发放、保障和优化运营能力;业务运营层,面向自治网络业务,提供客户、生态和合作伙伴的使能和运营能力。4个闭环实现层间全生命周期交互,其中包括:用户闭环实现上述3个层级之间和其他3个闭环间的交互;业务闭环实现业务和服务运营层之间的交互;服务闭环实现服务、网络和IT资源运营层之间的闭环;资源闭环实现以自治域为粒度的网络与信息通信技术(ICT)资源运营间的交互^[8]。

3GPP很早就开始了网络智能化的研究。R15版本首次引入网络数据分析功能(NWDAF),并将其作为AI网元。R17对NWDAF的架构和功能进行了增强,包括NWDAF的

▲图5 TM Forum的自治网络框架^[8]

逻辑功能拆分及逻辑功能间的交互，多NWDNF实例协作进行数据训练和模型共享，新功能实体的引入以提高数据收集效率，增强实时性等。R17还推进管理面的智能化，引入管理数据分析（MDA）。根据R18阶段发布版本，NWDNF进行了功能强化和解耦，将逻辑分析能力独立，并可调用机器学习（ML）模型和能力，通过树形级联调取其他NWDNF数据联合分析，将ML模型和能力集中进行统一管理。3GPP预计将在2025年下半年（R20）开始对6G技术进行标准化^[7]。

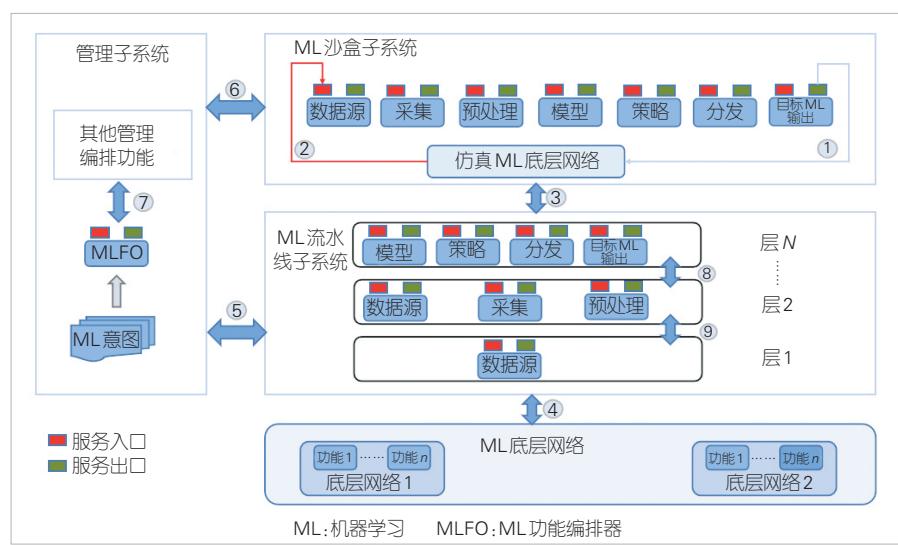
ITU-T SG13成立了未来网络机器学习焦点小组（ML5G），致力于ML在5G及未来网络中的应用。该小组完成了关于AI用例、架构框架、智能级别、数据处理、ML功能编排器（MLFO）、服务框架等多项技术规范，提出一套针对机器学习的管理子系统，如图6所示。针对机器学习全生命周期中各阶段所需的不同功能，该小组还提出了跨多域、多云、不同层级的多层次ML工作流。其中，完整的网络分析功能由一组ML管道节点构成，包括数据源、数据采集器、数据预处理器、ML模型、模型输出分配器和接收器等。这些节点可视为逻辑实体，具体部署位置由功能编排器来管理。为降低由于模型不确定性给网络带来的风险，该架构还采用了沙盒（sandbox）概念，作为独立环境专门用于ML模型的训练、测试和评估^[9]。

IMT-2030（6G）推进组于2019年6月由中国工业和信息化部等部委共同推动成立，是中国研究6G的一个组织，是聚合中国产学研用力量、推动中国6G技术研究和开展国际交流与合作的主要平台，对6G愿景、典型场景、无线技术和网络架构等展开系列研究。在IMT-2030网络组标准研讨中，形成了面向6G网络的智能内生网络体系框架雏形。该体系框架从逻辑层次上分为3层^[10]，从下到上依次为异构资源层、功能和编排管理层以及能力开放层，如图7所示。

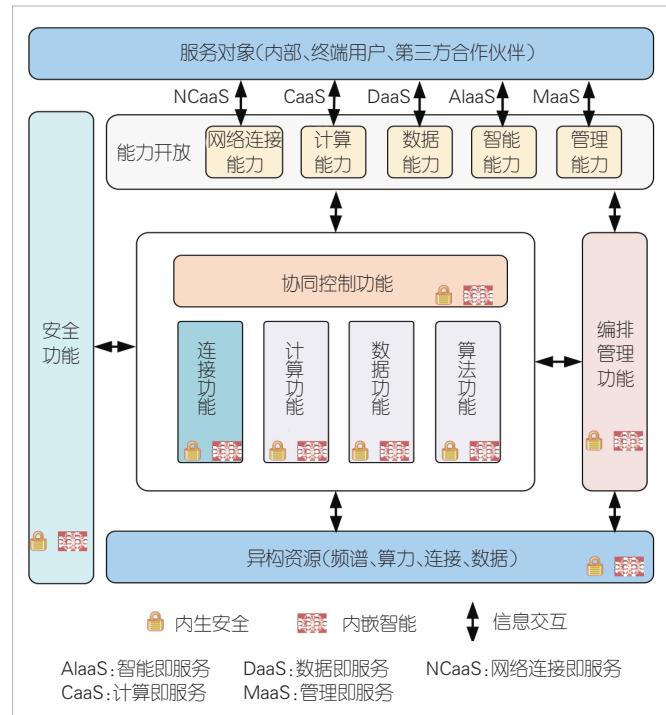
异构资源层提供网络连接、频谱、算力和数据等多维资源，是基础设施层；功能和编排管理层包括编排管理功能、协同控制功能、连接功能、计算功能、数据功能和算法功能，通过对多维资源（算力、算法、数据、连接）的实时监控、融合调度、联合编排，以及对各种任务实例的全生命周期的实时管理与控制，来提供AI相关服务，该层是提供AI服务的核心功能层；能力开放层作为服务的开放平台，以统一的服务化接口向任何潜在的服务消费者提供各类服务。

该体系框架的核心在于将网络连接与算力、算法、数据等AI三要素协同考虑，形成AI四要素，并通过对AI四要素的融合编排和协同调度，构建网络分布式智能服务的基础。同时，通过引入更加高效和实时的协同控制功能，在编排管理给出的编排流基础上，该体系实现AI四要素的实时协同管理和动态调整，与编排管理功能形成功能分层，协同实现低时延的高效网络管理，并实现跨区跨域间协同控制信息的高效交互，从而有效支持6G网络的智能内生。

综上所述，与学术界、产业界研究相对应，标准化工作



▲图6 TU-T ML5G架构^[9]



▲图7 IMT-2030网络组智能内生网络体系框架图^[10]

也包含智能能力生成、意图驱动的自智网络、AI要素协同、AI能力编排、支持分布式智能等相关方向。但目前的标准化工工作仍处于预研阶段，系统性设计智能内生网络功能及流程的工作还比较少，但相关工作也引发了业界的广泛关注，更具体的工作将进一步展开。

3 智能内生网络关键技术

面对规则式算法灵活性差、意图功能支持程度较低、AI单元协作能力较差、深度学习模型可解释性差等问题，6G智能内生网络在智能能力生成、智能化程度、AI能力部署、AI可信度等方面均面临诸多挑战。为应对这些挑战，本文从知识表征与构建、意图驱动、分布式AI和AI可解释性4个角度分析6G智能内生网络的关键技术。

3.1 知识表征与构建技术

知识将在6G智能内生网络中发挥重要的作用。知识不仅是网络数据，还可以是相关网络规律、机理、策略等的进一步凝练，具有历史记载、客观事实、经验积累、客观反馈等形式。知识除了人类所能总结的知识库、常识库外，还包括机器所能理解的知识。因此6G网络多维主客观知识的表征与构建是基础技术，包括知识获取、知识抽取、知识融合和知识存储等方面。

知识获取是获取不同来源、不同结构的数据，是知识构建的前提。结构化和半结构化数据通常包括网络状态数据、网络运行和故障日志、运维手册等，只需要简单读取预处理和映射即可作为后续数据分析系统的输入；而非结构化数据如专家经验、意图命令等则需要借助自然语言处理、信息抽取等技术来帮助提取有效信息。知识抽取是知识构建的核心，包括知识建模、实体抽取、属性抽取、关系抽取等，抽取方法与数据类型及结构化程度密切相关，常用技术包括知识映射、自然语言处理与深度学习等。知识融合是对不同来源、不同语言或不同结构的知识进行融合，从而对于已有知识进行补充、更新和去重，包括内容清洗、分块与记录连接等子任务^[11]。知识存储则综合考虑应用场景、系统性能、推理能力各方面需求，并使用图数据库或资源定义框架数据库进行存储，部分场景下也可使用关系型数据库进行存储。

3.2 意图驱动技术

6G智能内生网络将是意图驱动的网络。6G网络将以更高级别抽象的方式提取业务或用户意图，借助AI技术实现意图的识别、转译和验证，并在网络状态感知和精准预测的基础上，基于意图完成网络自动化部署配置、网络自主优化

和故障自愈等。意图驱动技术的应用，将完成网络全生命周期的自动化和智能化管理，极大地提升网络的运维效率，降低运维成本，提高对业务变化的响应速度。

意图驱动网络的关键技术包括意图转译和意图验证等。意图转译工作是意图驱动技术中的核心任务，它实现了用户意图到网络策略的转变。该工作主要采用自然语言处理的方法对意图进行处理，包括对用户意图进行关键字提取、词法分析、语义挖掘等操作，从而获得用户期望的网络运行状态。意图转译工作使用智能化的方法生成网络策略。在策略下发之前，策略的可执行性需要被验证，主要包括资源可用性、策略冲突及策略正确性3个方面。为实现资源可用性验证，网络状态信息数据库需要被维护；在策略冲突方面，意图功能模块需要检测待下发策略与网络当前策略是否存在冗余、覆盖、泛化、相关、重叠等冲突，并应通过设置优先级等方法进行冲突消解；在策略正确性验证方面，可采用形式化验证方法，将复杂系统建模为数学模型，开展模型检查、定理证明、符号执行以及符号执行问题（SAT）/可满足性模理论（SMT）求解等^[12]。

3.3 分布式AI技术

空天地海一体的6G网络是人、机、物融合的网络，将产生大量且多样的数据。这些数据分布在不同的网络、系统、网元上。如果将这些数据集中起来训练，将会产生高昂的计算及传输成本，并带来安全隐患，因此，分布式AI将是实现6G智能内生网络的关键技术。利用大规模分布式边缘设备的潜力，分布式AI技术可以解决集中式AI的瓶颈问题，降低通信开销并解决数据隐私保护问题。分布式AI和群智式的推理协同，将构筑6G全新的智能生态系统。

分布式AI技术主要分为分布式AI训练和分布式AI推理两方面。对于分布式AI训练问题，联邦学习是一种经典架构，参与训练的客户端无须上传本地数据，只需将训练后的模型参数更新上传，再由边缘服务器节点聚合、更新后下发给参与学习的客户端。空中计算技术可以利用无线链路上行多址接入信道的信号叠加特性，通过通信和计算的一体化设计，有效降低分布式训练过程中的通信开销和时延。分布式AI推理，即在网络边缘分布式执行AI模型。考虑到边缘节点的计算存储资源有限，如何减小及优化模型在分布式推理中显得尤为重要。常见模型压缩方法包括网络剪枝、知识蒸馏、参数量化、结构优化等^[13]。

3.4 AI可解释性技术

AI模型凭借其高精确度、强泛化性等重要优势，将与

6G网络深度融合，构筑智能内生网络。但是，AI模型特别是深度学习模型，参数多、结构复杂，人们往往难以对其决策进行预判和解释。AI模型的黑盒特性对其在6G网络中的部署提出了挑战，可解释性的缺乏将给系统可靠性带来很多潜在风险，引发道德和法律方面的问题。因此，AI可解释性技术对于建设可靠稳健安全可信的6G智能内生网络至关重要^[14]。

AI可解释性可以分为事前可解释性和事后可解释性。事前可解释性通过训练结构简单、可解释性好的模型，如朴素贝叶斯、线性回归、决策树、基于规则的模型等，或将可解释性结合到具体的模型结构中来实现模型本身的内置可解释性，如广义加性模型、自注意力机制等。事后可解释性指通过开发可解释性技术解释已训练好的AI/ML模型，分为全局可解释性和局部可解释性。全局可解释性旨在帮助理解复杂模型背后的整体逻辑及内部的工作机制，例如规则提取、模型蒸馏、激活最大化解释等；而局部可解释性旨在理解机器学习模型针对某一个输入样本的决策过程和决策依据，例如梯度反向传播解释、特征反演解释等^[15]。

4 结束语

面向6G实现万物智联、提供泛在智能的愿景，智能内生将成为6G网络的核心基因。本文提出了6G智能内生网络的概念和特征，分析了6G智能内生网络架构的学术研究和标准化进展，指出了6G智能内生网络的关键技术，为未来6G智能内生网络发展提供了参考。

目前6G研究方兴未艾，智能内生网络研究也处于热点阶段，但许多工作尚集中在概念、架构和功能设计及部分功能的验证上，系统性的工作还较少，融合AI和网络的标准化工工作也还在研究和推进中。6G智能内生网络仍面临高效性、安全性、可解释性等诸多方面的挑战，需要业界进一步的研究和推进。

致谢

感谢北京邮电大学张俊也博士对本文的贡献！

参考文献

- [1] 中国移动通信集团有限公司. 中国移动6G网络架构技术白皮书 [R]. 2022
- [2] 中国电信研究院, 中兴通讯. 6G网络架构展望白皮书 [R]. 2023
- [3] ZHOU F Q, LI W J, YANG Y, et al. Intelligence-endogenous networks: innovative network paradigm for 6G [J]. IEEE wireless communications, 2022, 29(1): 40–47. DOI: 10.1109/MWC.004.00320
- [4] 金哲, 张引, 吴飞, 等. 数据驱动与知识引导结合下人工智能算法模型 [J]. 电子与信息学报, 2023, 45(7): 2580–2594
- [5] 杨静雅, 唐晓刚, 周一青, 等. 意图抽象与知识联合驱动的6G内生智能网络架构 [J]. 通信学报, 2023, 44(2): 12–26. DOI: 10.11959/j.issn.1000-436x.2023016
- [6] 吴建军, 邓娟, 彭程晖, 等. 任务为中心的6G网络AI架构 [J]. 无线电通信技术, 2022, 48(4): 599–613. DOI: 10.3969/j.issn.1003-3114.2022.04.005
- [7] 亚信科技, 清华大学, 中国电信, 等. 6G OSS技术白皮书 [R]. 2023
- [8] TM Forum. 自智网络白皮书3.0 [R]. 2021
- [9] Architectural framework for machine learning in future networks including IMT-2020: ITU-T Y.3172 [S]. 2020
- [10] IMT-2030(6G)网络技术工作组. 面向6G网络的智能内生体系架构研究 [R]. 2022
- [11] 6GAN. B5G/6G网络智能数据分析: 网络数据采集、知识表示与推理、特征数据集构建与评估 [R]. 2022
- [12] 李福亮, 范广宇, 王兴伟, 等. 基于意图的网络研究综述 [J]. 软件学报, 2020, 31(8): 2574–2587. DOI: 10.13328/j.cnki.jos.006088
- [13] SHI Y M, YANG K, JIANG T, et al. Communication-efficient edge AI: algorithms and systems [J]. IEEE communications surveys & tutorials, 2020, 22(4): 2167–2191. DOI: 10.1109/COMST.2020.3007787
- [14] WU Y L, LIN G Z, GE J G. Knowledge-powered explainable artificial intelligence for network automation toward 6G [J]. IEEE network, 2022, 36 (3): 16–23. DOI: 10.1109/MNET.005.2100541
- [15] 纪守领, 李进锋, 杜天宇, 等. 机器学习模型可解释性方法、应用与安全研究综述 [J]. 计算机研究与发展, 2019, 56(10): 2071–2096. DOI: 10.7544/issn1000-1239.2019.20190540

作者简介



李文璟，北京邮电大学教授、博士生导师，中国通信学会高级会员；主要研究领域为无线网络智能管理、B5G/6G网络架构；先后主持国家“863”计划课题、国家科技重大专项项目、国家重点研发计划项目、国家自然科学基金重点项目及面上项目等10余项；出版论著2本，以第一起草通信行业标准20余项。



喻鹏，北京邮电大学未来学院副院长、副教授、博士生导师，IEEE/EAI高级会员、中国通信学会高级会员；主要研究方向为B5G/6G网络管理与优化；近年来主持/参与了国家级项目10余项，参与起草了国际行业/企业标准10余项，荣获科技奖励5次、国际期刊会议最佳论文奖4次；发表论文60余篇。



张平，中国工程院院士，北京邮电大学教授、博士生导师、网络与交换技术全国重点实验室主任，国家自然科学基金委“创新研究群体”带头人；主要研究方向为先进移动通信系统等；先后获国家科学技术进步奖特等奖1项、国家科学技术进步一等奖1项、国家技术发明奖二等奖3项、国家科学技术进步奖二等奖2项，获首届“全国创新争先奖”奖章、光华工程科技奖、何梁何利基金科学与技术进步奖，带领的团队入选首批“全国高校黄大年式教师团队”。



面向6G的星地融合网络架构

Integrated Satellite and Terrestrial Network Architecture for 6G

徐晖/XU Hui^{1,2}, 陈山枝/CHEN Shanzhi²,
艾明/AI Ming^{1,2}

(1. 大唐移动通信设备有限公司, 中国北京 100083;
2. 无线移动通信全国重点实验室(中国信息通信科技集团有限公司), 中国北京 100019)
(1. Datang Mobile Communications Equipment Co., Ltd, Beijing 100083, China;
2. State Key Laboratory of Wireless Mobile Communications (CICT), Beijing 100019, China)

DOI: 10.12142/ZTETJ.202305003

网络出版地址: <http://kns.cnki.net/kcms/detail/34.1228.TN.20231017.0948.002.html>

网络出版日期: 2023-10-18

收稿日期: 2023-08-10

摘要:通过对卫星星座的发展和星地融合网络研究的回顾,明确了6G的星地融合网络的发展趋势,提出了智简赋能的6G网络体系架构和弹性可重构的6G星地融合架构,并分析了星地融合网络中的关键技术问题,包括星上轻量化虚拟化技术、星上边缘计算功能以及广播/多播技术。认为6G星地融合网络将通过星地协同实现网络资源和计算资源的统一调度,同时可以根据业务需求和网络状态智能实现网络功能的按需弹性部署。

关键词: 6G网络架构; 星地融合网络架构; 边缘计算; 广播/多播

Abstract: Through a review of the development of satellite constellations and research on the integrated satellite and terrestrial networks, the trend of the integrated satellite and terrestrial networks for 6G is put forward. An intelligent simplicity empowerment 6G network architecture and a resilient and reconfigurable integrated satellite and terrestrial architecture are proposed. Then the key technical issues in the integrated satellite and terrestrial networks, including lightweight virtualization technology onboard satellites, onboard edge computing function, and broadcast/multicast technology are analyzed. It is believed that 6G integrated satellite and terrestrial networks will achieve unified scheduling of network and computing resources through satellite-terrestrial cooperation and can intelligently deploy network functions flexibly according to service requirements and network conditions.

Keywords: 6G network architecture; integrated satellite and terrestrial network architecture; edge computing; broadcast/multicast

引用格式: 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, 29(5): 9–15. DOI:10.12142/ZTETJ.202305003

Citation: XU H, CHEN S Z, AI M. Integrated satellite and terrestrial network architecture for 6G [J]. ZTE technology journal, 2023, 29(5): 9–15. DOI:10.12142/ZTETJ.202305003

随着信息和通信技术的发展,6G网络将为用户提供丰富沉浸式体验,通过协同和融合的方式实现无处不在的连接。6G网络将解决连接、覆盖、容量、数据速率和终端移动性方面的问题,实现全域覆盖,为用户提供获得教育、卫生、农业、运输、物流等服务。同时,6G网络将通过地面蜂窝网络与非地面蜂窝网络的互通和融合,支持全域的服务连续性。卫星通信网络作为地面移动通信网络的重要补充,在6G网络中将与地面移动通信网络深度融合。6G网

络将通过星地融合真正实现无处不在的连接,弥合数字鸿沟^[1-2]。

1 6G星地融合网络的愿景和需求

1.1 6G网络愿景

5G大规模的商用,提升了用户的通信服务体验,为垂直行业的数字化转型提供了技术手段。面向2030及未来,6G网络通过人机物智能互联、协同共生,满足经济社会高质量发展需求,服务智慧化生产与生活,推动并构建普惠智能的人类社会^[3]。

基金项目:国家重点研发计划项目(2020YFB1807901)

作为负责全球移动通信标准化的组织，国际电信联盟（ITU）在2023年6月发布了《IMT面向2030及未来发展的框架和总体目标建议书》。该建议书从分析社会发展愿景、典型场景和应用需求入手，提出了6G网络的七大目标，即包容性、泛在连接、可持续性、创新性、安全性/隐私性/弹性，以及标准化和互操作、互通性。

6G对5G的三大典型场景进行了增强和扩展，提出了沉浸式通信、超大规模连接、极高可靠低时延、人工智能与通信的融合、感知与通信的融合、泛在连接六大场景。6G将实现人、机、物的智能连接，实现物理世界和虚拟世界的连接^[4]。6G业务将呈现出泛在智能、泛在计算、沉浸式多媒体和多感官通信、数字孪生和虚拟世界、智能工业应用、数字健康与福祉、泛在连接、传感和通信融合、可持续性九大趋势。

6G将提供泛在连接能力，解决目前没有覆盖或者几乎没有覆盖的地区如海洋、沙漠等的通信问题，加强连接，缩小数字鸿沟。6G将通过统一的通信体制和网络实现星地的全球无缝覆盖，实现以用户为中心的网络^[5]，满足用户随时随地的通信需求。

1.2 星地融合网络需求

星地融合是6G网络的一个主要特征，通过高轨卫星网络、中低轨卫星网络和地面移动通信网络等共同组成立体覆盖移动通信网络，从体制、协议、网络、业务、终端等方面实现天基网络和地面移动通信网络的深度融合，满足地面和立体空间的全域、全天候的泛在覆盖需求，从而实现用户随时随地按需接入^[6]。6G网络通过星地融合网络的多接入融合，不仅为用户业务提供较好的服务质量（QoS）和较高的可靠性，还为用户提供可靠、连续的通信服务和更高的用户体验^[7-8]。

星地融合网络要考虑卫星通信拓扑结构的动态性、复杂性和可扩展性，以及空间节点的高速移动性和有限的存储及处理能力等特点^[9]，通过融合的网络架构实现统一高效的智能网络，促进6G的多元信息传输的共享，提升网络传输效率和性能，降低建网与维护成本，灵活部署网络，拓展6G业务应用范围。

传统的卫星通信网络与地面移动通信网络各自独立发展，独立组网，通过网关实现卫星与地面网络业务互联互通，因此网络功能单一，资源无法统一协调；并且网络效率低，系统间互通性差，因此难以满足信息融合和综合利用的需求。

随着卫星通信和地面移动通信技术的不断发展，为了给

用户提供更好的业务体验，5G移动通信系统开始将卫星接入体制融合到移动通信系统中，即卫星通信系统采用与地面移动通信系统相同或者相似的通信体制。用户通过卫星通信系统接入到移动通信核心网中，实现统一管理，从而实现终端在卫星通信系统与地面移动通信系统之间的业务连续性。6G网络将融合多种空间平台，包括同步卫星、中低轨卫星、平流层浮空器以及飞机等，使星地构成一个统一的系统。整个系统的接入点、频率、接入网、核心网完全实现统一规划和设计，并采用协同资源调度实现一致的服务质量、星地无缝切换^[6]，为用户提供无感知的高质量的服务。

面向6G的星地融合网络将采用统一的空口体制和网络架构，实现统一的资源调度和管理、统一的终端标识、统一的业务管理和调度、统一的网络平台和频率的共享共用，以及统一管理^[10]。

2 星地融合网络研究现状

2.1 卫星星座的发展现状

随着互联网业务和宽带通信技术的不断发展，卫星通信正在朝着宽带化和网络化发展。同时，批量化的卫星制造能力降低了卫星的成本，这使得巨型星座的发展成为可能。新型的卫星互联网星座已经扩展至上万颗卫星，如Starlink、OneWeb、Lightspeed和Kuiper等。这些星座利用Ku/Ka/Q/V频段点波束实现全球覆盖，可以满足未来业务增长的需求^[11]。

在组网方面，星间组网已成为发展趋势，卫星互联网星座可以通过星间链路实现全球无缝连接。由于低轨卫星星座在覆盖范围、网络时延、系统容量等方面能力具有较为明显的优势，新兴的卫星互联网星座大多采用中低轨道，6G网络可以通过低轨卫星星座构建低时延的星地融合网络系统。与低轨卫星相比，高轨卫星在覆盖范围、网络复杂性和移动性管理等方面具有一些较大的优势，因此6G星地融合网络需要考虑到高、低轨卫星的融合组网，从而可以实现优势互补。

在载荷能力方面，半导体技术的进步使卫星具有更强的星上处理和存储能力，而激光星间链路技术大幅提升了星间通信速率。卫星通信载荷朝着轻量化、软件化和多样化的方向发展。目前大部分卫星星座具备星上处理能力，可以对接收的信令或者数据进行解析、存储和转发。同时，星上载荷可以支持在轨功能重构和自适应调整。卫星与卫星之间可以通过星间链路实现数据或信令的转发。卫星的星上处理能力和星间链路的传输能力使得星地融合组网成为可能，这增强

了系统的灵活性。星地融合网络将实现通信和计算融合，构建星上网络基础设施，实现星上网络功能和算力资源的弹性部署与调度。卫星星座通过与地面核心网以及终端的协同，形成云-边-端协同的星地融合网络。

2.2 卫星通信与5G的融合组网标准进展

国际上，ITU、第3代合作伙伴计划（3GPP）等主要标准化组织已经启动了卫星通信与5G的融合组网研究工作。2019年ITU-R第4研究组第2工作组提出了星地融合的4种应用场景，包括中继到站、小区回传、动中通及混合多播场景，并提出支持这些场景必须考虑的关键因素，包括多播/单播支持，智能路由、动态缓存管理及自适应流的支持，服务质量保证，网络功能虚拟化/软件定义网络兼容性等^[12]。在2023年6月召开的国际电信联盟无线电局卫星研究组第2工作组（ITU-R SG4 WP4B）全会通过了由中国提出的卫星国际移动通信（IMT）未来技术趋势的研究报告立项。该报告研究面向IMT-2030的卫星应用的驱动力以及关键技术，包括手机直连卫星通信、星上处理、星间链路、高低轨卫星协同、星地频谱共享技术等，同时也研究星地融合网络的部署/架构、终端、隐私和安全等内容。

在网络架构及关键技术方向上，ITU-T SG13工作组从2020年开始开展面向IMT-2020网络的固定、移动与卫星融合的需求、架构、移动性管理、接入管理、边缘计算、本地数据转发、策略管理等方向的标准化工作。目前ITU-T SG13已经开始讨论下一个工作周期的研究范围和内容，其中星地融合网络将是ITU-T SG13工作组的一个重要的标准化方向。

3GPP在5G系统中将卫星作为一种接入方式。从REL 16开始，3GPP在网络架构和无线接入两个方向进行卫星与5G融合的研究和标准化工作。在REL 16和REL 17中，3GPP主要针对透明转发的组网场景进行研究和标准化工作，主要包括支持卫星接入的5G接入网架构、物理层、空口协议、移动性管理机制、馈电切换、服务质量控制机制等体制增强以及支持卫星监管要求的5G系统增强。在REL 18中，3GPP主要针对卫星覆盖增强、10 GHz以上频段新空口非地面通信（NR-NTN）网络部署、网络验证终端位置、星地网络和星星网络之间的移动性管理和业务连续性增强、非连续卫星覆盖的增强、卫星回传增强和用户面功能上星等进行了研究和标准化工作。在REL 19中，3GPP SA1继续研究星地融合的场景和需求，在卫星的存储转发、与用户设备（UE）通信、全球导航卫星系统（GNSS）无关操作、定位增强等方面进行场景和需求的研究和标准化工作。

3 6G星地融合网络架构

3.1 6G网络体系架构

业务场景、商业模式需求的多样化，以及社会责任的持续性，这些都使6G网络面临新的挑战。6G要突破从1G到5G的传统陆地移动通信系统，实现星地融合、全域覆盖；要突破从1G到5G的传统蜂窝通信架构，朝着智简方向发展；要突破从1G到5G中只提供通信服务的方式，为业务提供更多的网络内生能力。

随着网络接入规模的持续增大和网络需求的多样化发展，网络变得越来越复杂。为了避免网络复杂叠加，6G网络需要设计智简的架构，以提高网络运行效率，降低运维成本。同时，为了满足多样化场景的业务需求，6G网络将突破中心化的限制，构建分布式自治的网络，实现以用户为中心的控制和管理，满足天、地、空、海等多种异构接入场景和网络性能需求。随着网络规模的不断增大，6G网络呈现出集中加分布式联合部署的方式，实现网络资源互补和按需组网的目标。

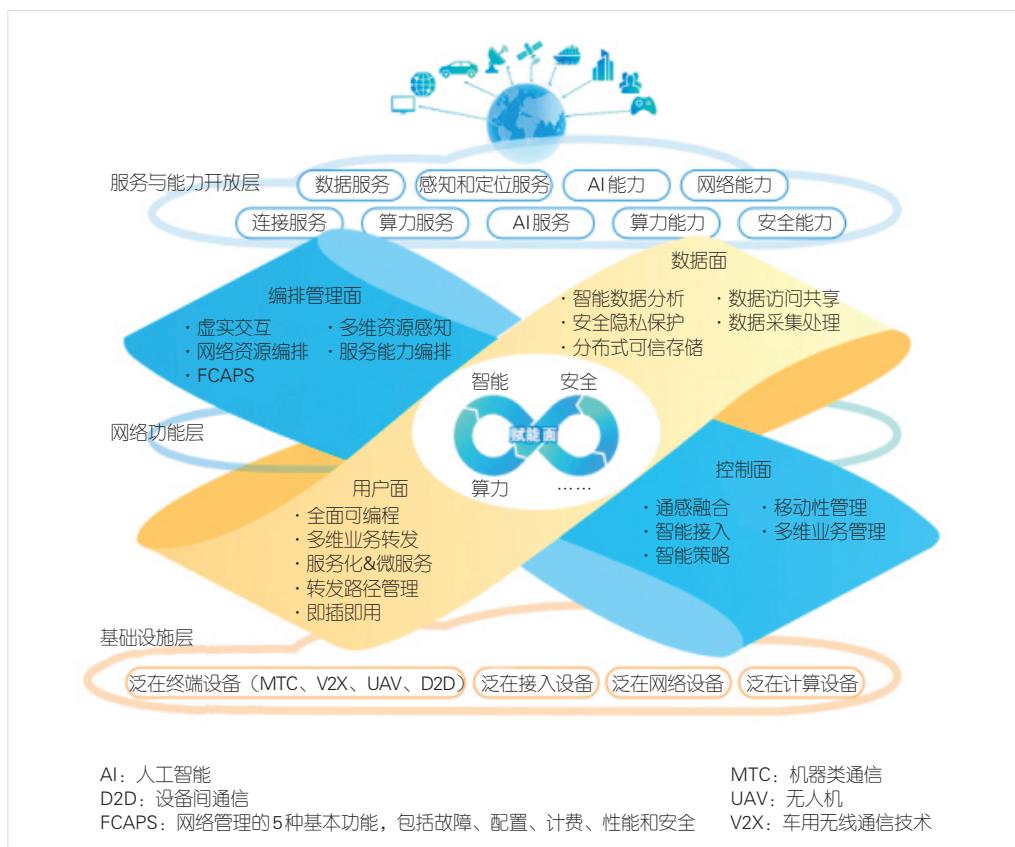
目前，移动通信网络与人工智能缺乏深度融合，与算力网络独立管理，网络安全防护方式缺乏动态性和流动性。6G网络架构需要融合智能内生、算力内生、安全内生的能力，实现网络无所不达，算力无处不在，智能无所不及。

5G网络基于服务化的架构使得控制面与用户面的功能分离。但5G的用户面和接入网服务化程度不高，网络全场景适应能力有限，这约束了网络灵活部署的能力。因此6G网络将进一步增强服务能力，实现软硬资源充分解耦，灵活调配，从而增加网络柔性，提升网络的适应能力。6G网络将通过可编程技术实现网络按需编排、资源灵活部署、路由动态规划、网络故障快速定位与修复，满足业务多样化的需求。同时，智能内生与可编程的深度融合，为6G网络的编排与实现提供了帮助。

6G网络采用智慧内生和高效简洁的轻量化网络架构，基于内生智能，按需编排网络功能和资源，从而降低网络复杂性。考虑到全场景的泛在连接、业务需求动态变化以及网络的灵活扩展，6G网络需要具备统一架构下的即插即用功能。同时6G网络是集连接、智能、算力、安全、数据等为一体的融合网络，需要支持将多种能力和服务进行一体化的编排管控。图1为智简赋能的6G网络体系架构。

6G网络架构的智简化体现在以下几个方面：

- 架构智简。6G网络支持服务化架构，可以面向场景需求灵活拆分、重组网络功能，通过对多维资源的统一编排，



▲图1 智简赋能的6G网络体系架构

实现以用户为中心的调度和管控, 建立弹性可重构的网络架构。

- 功能智简。网络功能将进行更细粒度的分割, 支持柔性按需编排原子化功能和网络功能的优化重构。

- 协议智简。全域采用统一协议支持星地融合跨域交互, 通过面向场景的协议可编程支持各行业差异化需求, 通过统一空口传输协议和组网协议提升端到端网络性能。

- 终端智简。终端支持星地融合等多种空口技术的融合, 实现终端无差别的接入。

- 基础设施智简。6G网络通过泛在资源池为网络提供可编程和动态的硬件或软件资源。

6G网络除了提供连接服务外, 还要为用户提供感知与定位服务、人工智能(AI)服务、数据服务、算力服务等服务。因此, 6G网络应具备极致弹性、灵活能力扩展的赋能能力, 为6G网络赋予所需的内生能

力, 从而根据网络和业务需求灵活构建网络服务。6G网络的赋能将体现在以下两个方面:

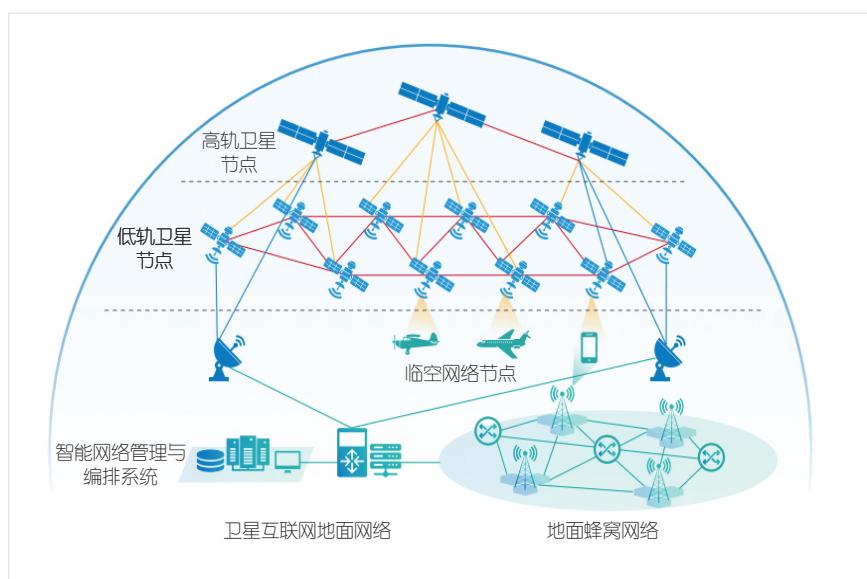
- 对内赋能: 通过原生网络功能原子化, 随需编排智能、算力、安全等功能, 赋予网络具备智能、算力、安全等内生能力。

- 对外赋能: 通过设计原生开放的机制, 为行业用户和第三方提供了可信并且随需的网络能力和服务等, 赋能用户和第三方智能、算力、安全等千人千面的相关需求。

3.2 6G星地融合网络架构

6G星地融合通信网络由地面移动通信网络、临近空间网络和天基网络组成, 是具有大时空尺度的异构网络,

如图2所示。星地融合网络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, 满足用户的无时不在的通信需要。



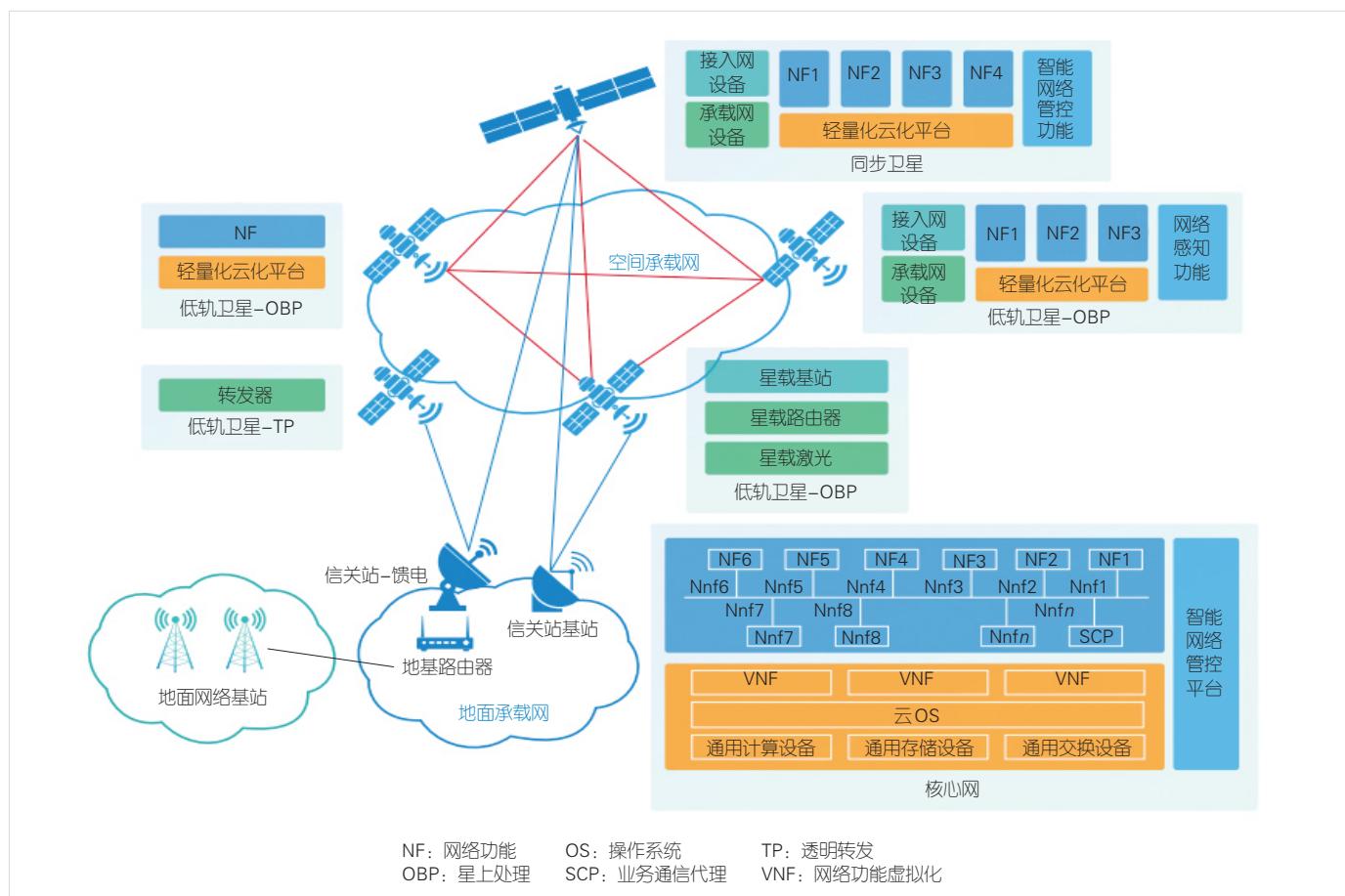
▲图2 星地融合立体网络

随着星上处理能力的不断提高，天基网络可以在星上实现信号处理、路由交换和信令处理。星地融合网络的运行环境与地面网络运行环境相比存在着较大的差异，因此星地融合网络需要针对不同的传输条件和传输时延并通过网络功能的柔性编排实现智简网络的灵活部署和优化。星地融合网络应具有简洁、敏捷、开放和资源随选等特点，因此需要尽量减少网络层级和接口数量，降低运维的复杂性。同时，星地融合网络的弹性可重构能力可以实现网络和业务的按需快速部署，通过网络资源的统一编排和管理，可以实现端到端的资源统一管理。星地融合弹性可重构的网络架构如图3所示。

星地融合网络采用服务化的架构，可以根据业务和组网需求进行网络功能的按需部署，并可将其在卫星和地面网络间进行柔性分割。例如，可以在地面部署完整的移动通信核心网和接入网，在卫星上根据组网需求和卫星能力进行网络功能的分割和灵活部署，提供定制化网络服务能力，从而提升网络的全场景适应能力。由于星上资源有限，为了实现网

络功能的柔性部署，需要在星上部署轻量化的虚拟化资源，提高不同/异构节点上资源的通用性，实现星上功能的重构；也可以通过容器等虚拟化技术在星上部署定制化的网络功能，降低部署网络功能的资源开销；还可以对网元功能进行轻量化设计和细粒度分割，并根据组网形态、业务需求、网络资源情况和节点处理能力，在卫星和地面节点之间合理部署网元功能，实现网络功能的定制和按需重构。

由于星地融合网络具有全域覆盖、随遇接入和动态变化的特点，因此6G网络需要对异构网络的资源进行统一和智能的管理。为此，可以在卫星节点和地面部署网络编排系统。由于卫星节点资源受限，因此卫星上部署的编排器采用了轻量化的网络编排技术，负责感知卫星节点和链路的状态，并可以根据地面智能网络编排系统的指令以及卫星节点的可用性、安全性、负载等进行网络重构。将在地面部署的智能网络编排系统，负责星地融合网络资源的统一管理。智能编排系统将根据用户需求数据、网络运行数据、网络拓扑数据、星历数据等，通过AI算法分析与业务需求匹配的网



▲图3 星地融合网络架构

络能力，实现按需、自动化地对网络资源进行管理和编排；智能网络编排平台还具备端到端的网络策略管理能力，结合用户的业务需求、网络状态和资源现状，进行智能决策与策略下发，从而可以实现网络功能和性能的自动化编排与管理。

3.3 星地融合网络的边缘计算

随着新业务的不断涌现，作为远程地面数据中心的中继，卫星网络已无法继续提供高效的数据处理响应。因此星地融合通信网络需要考虑边缘计算功能，在星地之间进行柔性分割边缘，把全局性强、复杂度高、时延低的业务放在地面网络平台上处理，把实时性要求高、计算复杂度低的业务放在卫星节点上处理，通过在轨计算服务，降低时延并实现高效的业务分发。例如，利用卫星预先数据缓存、卫星广播/组播功能可以大大减少星地融合通信网络的前向链路上的流量；将部分卫星数据（如地球图像和天气观测）的处理转移到卫星上，可以大大地降低需要回传的数据量，节省馈电链路的带宽。

6G分布式网络可以在边缘的云化基础设施上部署边缘计算节点，实现对第三方应用的部署和管理。对于星地融合网络来说，卫星节点可以作为分布式的边缘节点来部署边缘计算平台，与部署在地面的边缘计算节点协同，实现跨域的多级边缘计算协同。

随着星上电子器件的不断发展，星上的计算能力有较大的提高，但仍需要网络在云、网、边间按需分配，灵活调度计算资源、存储资源以及网络资源。通过边缘计算智能协同机制，智能网络编排系统可以根据网络状态和算力状态，为用户的业务需求匹配最佳的计算资源，并计算从用户到目标算力节点间的最优路径，确定业务的转发路径。星上部署的边缘计算节点可以利用相对有限的算力资源和网络资源执行部分的边缘算力任务，而地面的中心算力节点具有丰富的算力资源和网络资源，是终端和边缘的有力支撑，可以完成边缘节点无法完成的复杂算力任务。

3.4 广播和多播技术

卫星网络通信系统本身具有覆盖面积大、接入速度快、高带宽等特性，易于实现广播和组播业务，是中国广播电视台传输覆盖的重要手段和渠道，更好地满足人民群众日益增长的精神文化需求。《广播电视台和网络视听“十四五”科技发展规划》中指出，在“十四五”期间，要建设交互卫星广播电视台系统，构建双向、互动、跨网络的新型卫星直播融合业务平台，提升直播卫星业务承载能力，探索直播卫星与宽带

通信卫星、卫星互联网的融合应用。因此星地融合网络可以通过广播和多播技术为用户提供广域的信息内容服务、公共安全和应急任务、集团客户的群组通信等。

在接入网技术方面，为了提高系统资源的效率和用户的服务体验，星地融合网络需要解决不同卫星和不同小区之间业务连续性问题^[13-14]。星地融合网络可以使用单一频点来提供广播服务，以便空闲态终端和连接态终端都可以接收到广播业务，从而降低网络开销。星地融合网络可以根据广播区域的要求，通知特定的卫星或波束向特定的终端发送广播信息。同时终端可以对信息内容进行反馈，提高传输的可靠性。对于组播业务来说，星地融合网络需要解决高效的组播管理和组播路由问题。根据用户和业务需求为特定的终端建立组会话，可以实现组内用户的相互通信。当同一组播组的用户在同一波位内时，只需要在无线空口传输一份信息，便可大大节省资源开销；当同一组播组的用户在同一卫星下的不同波位时，需要在不同波位上分别发送组播信息；当同一组播组的用户在不同卫星下时，需要在相关卫星的相应波位上分别发送组播信息，并需要网络支持组播路由。对于低轨道卫星/中轨卫星通信网络来说，同一组播组的用户通常处在不同卫星下，并频繁地进行星间切换，因此需要高效的组播管理机制和组播路由协议。

在网络技术方面，可以采用移动通信的广播/多播网络架构进行组网；星上基站应支持广播/多播业务；核心网网络功能部署在地面，实现广播/多播业务的授权、会话管理及用户面管理等功能，从而实现热点区域、热点内容的推送与广播，为广域分散的移动用户提供群组通信、多方视频会议、双向对讲、视频点播等新型互联网应用。随着星载边缘计算技术的不断成熟，广播/多播技术可以和边缘计算融合，通过在星上部署支持广播/多播的用户面功能和边缘应用服务器，并通过在星上边缘云节点缓存热点内容向用户按需分发，实现定时广播、分区广播、自由点播等卫星广播/多播增值业务。

4 结束语

随着无线通信技术的发展，星地融合通信已经成为地面移动通信和卫星通信发展的必然趋势。星地融合网络将完成从卫星通信和地面通信的标准体制兼容到卫星通信和地面的系统全面融合的演进，并将成为6G网络区别于传统的地面移动通信系统的重要特征。6G星地融合网络在继承了6G网络新架构和新特征的同时，需要根据卫星的特点进行技术优化和改进，真正实现全域覆盖，万物智联。

参考文献

- [1] ITU-R. Framework and overall objectives of the future development of IMT for 2030 and beyond [R]. 2023
- [2] CHEN S Z, LIANG Y C, SUN S H, et al. Vision, requirements, and technology trend of 6G: how to tackle the challenges of system coverage, capacity, user data-rate and movement speed [J]. IEEE wireless communications, 2020, 27(2): 218–228. DOI: 10.1109/MWC.001.1900333
- [3] IMT-2030(6G)推进组. 6G总体愿景与潜在关键技术白皮书 [R]. 2021
- [4] 彭健, 孙美玉, 滕学强. 6G愿景及应用场景展望 [J]. 中国工业和信息化, 2020 (9):18–25. DOI: 10.3969/j.issn.1674–9138.2020.09.003
- [5] CHEN S Z, CHEN L, HU B, et al. User-centric access network (UCAN) for 6G: motivation, concept, challenges and key technologies [EB/OL]. [2023–08–10]. <https://ieeexplore.ieee.org/abstract/document/10148944>
- [6] 陈山枝. 关于低轨卫星通信的分析及我国的发展建议 [J]. 电信科学, 2020, 36 (6): 1–13. DOI: 10.11959/j.issn.1000–0801.2020181
- [7] CHEN S Z, SUN S H, KANG S L, et al. System integration of terrestrial mobile communication and satellite communication—the trends, challenges and key technologies in B5G and 6G [J]. China communications, 2020, 17(12): 156–171. DOI: 10.23919/JCC.2020.12.011
- [8] 孙韶辉, 戴翠琴, 徐晖, 等. 面向6G的星地融合一体化组网研究 [J]. 重庆邮电大学学报: 自然科学版, 2021, 33(6): 891–901
- [9] 徐晖, 孙韶辉. 面向6G的天地一体化信息网络架构研究 [J]. 天地一体化信息网络, 2021, 2(4):1–9. DOI: 10.11959/j.issn.2096–8930.2021037
- [10] 徐晖, 缪德山, 康绍莉, 等. 面向天地融合的卫星网络架构和传输关键技术 [J]. 天地一体化信息网络, 2020, 1(2):1–10
- [11] 陈东, 仲小清, 邓恒, 等. 宽带卫星通信网络技术发展态势与发展建议 [J]. 前瞻科技, 2022, 1(1): 86–93
- [12] 康绍莉, 缪德山, 索士强, 等. 面向6G的空天地一体化系统设计和关键技术 [J]. 信息通信技术与政策, 2022(9): 18–26
- [13] CHEN S Z, SUN S H, XU G X, et al. Beam-space multiplexing: practice, theory, and trends, from 4G TD-LTE, 5G, to 6G and beyond [J]. IEEE wireless communications, 2020, 27(2): 162–172. DOI: 10.1109/mwc.001.1900307
- [14] CHEN S Z, SUN S H, MIAO D S, et al. The trends, challenges and key technologies of beam-space multiplexing in the integrated terrestrial-satellite communication for B5G and 6G [J]. IEEE wireless communications, 2022, 29: 1–10. DOI: 10.1109/MWC.004.2200085

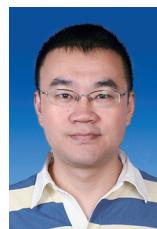
作者简介



徐晖, 大唐移动通信设备有限公司创新中心技术总监; 主要研究方向为移动通信网络、移动通信安全、星地融合网络等; 先后主持和参与国家“863”计划项目、国家重大专项和国家重点研发计划近10项; 已发表论文10余篇。



陈山枝, 中国信息通信科技集团有限公司副总经理、总工程师、科技委主任, 无线移动通信全国重点实验室主任, 教授级高工, 国家杰出青年科学基金获得者; 长期从事移动通信关键技术研发、国际标准研制和产业化工作, 主要研究方向为车联网、星地融合移动通信、B5G/6G; 曾主持国家“863”计划项目、国家重大专项和国家重点研发计划等10余项; 发表论文90余篇, 出版学术专著6部, 授权发明专利80余项。



艾明, 大唐移动通信设备有限公司创新中心标准项目经理, 曾担任3GPP TSG CT副主席; 主要研究方向为移动通信网络、网络人工智能、星地融合网络等领域; 先后负责多个3GPP国际标准制定项目, 主持国家重大专项和国家重点研发计划各1项。

6G 网络架构展望



Prospects for 6G Network Architecture

刘玉芹/LIU Yuqin, 邢燕霞/XING Yanxia,
陈鹏/CHEN Peng

(中国电信股份有限公司研究院, 中国 广州 510630)
(Research Institute of China Telecom Co., Ltd, Guangzhou 510630, China)

DOI:10.12142/ZTETJ.202305004

网络出版地址: <http://kns.cnki.net/kcms/detail/34.1228.TN.20231016.1553.014.html>

网络出版日期: 2023-10-17

收稿日期: 2023-08-23

摘要: 从系统架构、网络功能、网络组网3个层面对6G网络架构进行阐述。在网络系统层面,从全局角度描述6G各层各面的关系,提出“三层四面”系统架构;在网络功能层面,从网络功能视图的角度描述6G功能服务的划分和组成,提出至简功能架构;在组网层面,从网络部署视图的角度描述6G网络之间的连接关系和组网形态,提出分层分布式组网架构。所提出的6G网络架构能够满足新业务新场景需求,降低网络复杂度,提升网络灵活性。

关键词: 6G; 网络架构; 网络简化; 分层分布式架构

Abstract: The network architecture is expounded from three aspects of the overall system, network function, and networking. From the aspect of the network system, the "three-layer and four-plane" system architecture is proposed by describing the relationships between various layers and planes of 6G from a global view; from the aspect of network function, the simplified function architecture is proposed by describing the division and composition of 6G function services from a network function view; from the aspect of networking, the hierarchical and distributed network deployment architecture is proposed by describing the connection relationships and network forms between 6G networks from the network deployment view. The proposed 6G network architecture can meet the requirements of new services and new scenarios, reduce network complexity, and improve network flexible.

Keywords: 6G; network architecture; network simplification; hierarchical and distributed architecture

引用格式: 刘玉芹, 邢燕霞, 陈鹏. 6G 网络架构展望 [J]. 中兴通讯技术, 2023, 29(5): 16–20. DOI:10.12142/ZTETJ.202305004

Citation: LIU Y Q, XING Y X, CHEN P. Prospects for 6G network architecture [J]. ZTE technology journal, 2023, 29(5): 16–20. DOI:10.12142/ZTETJ.202305004

当前全球6G处于统一愿景、形成技术框架的关键时期,国际电信联盟无线电通信部门5D工作组(IITU-R WP5D)于2023年6月完成了《IMT面向2030及未来发展的框架和总体目标建议书》^[1],汇集了全球6G愿景共识,总结了6G用户和应用的九大趋势,并提出了沉浸式通信、超可靠低延迟通信、大规模通信、泛在连接、人工智能与通信融合、感知与通信融合六大场景。除了要满足传统5G网络的3类场景需求外,未来6G网络还需要进一步提供泛在连接、通信与智能融合、通信与感知融合等服务。在技术指标方面,未来6G网络除了在传统性能指标方面有进一步的增强外,还提出了更多的新功能及性能指标。

此外,6G网络将应用于多样化的使用场景,如面向企业(2B)类的智能工厂,面向消费者(2C)类的多接入,

面向个人类的个人车内通信、个人及家庭数字医护等场景。不同的场景对网络的功能及性能需求也有所差异,同时对网络的灵活性、定制化也提出了更高要求^[2-6]。

后续6G将从愿景阶段进入架构与关键技术方案的研究和标准化阶段。网络架构设计是6G网络愿景落地的重要环节和技术手段,是未来网络部署的依据,也直接体现了6G网络的商业价值。当前各国研究机构都在进行网络架构及关键技术研究,包括美国的Next G^[7]、欧盟“Hexa-X”^[8-9]、5G PPP^[10]等。

根据上述分析,6G应用和场景对网络提出了新的、差异化功能和性能需求,同时对网络的灵活性也提出了更高要求。但当前网络在本质上还是集中式的架构,且结构复杂,面临着灵活扩展的挑战。另外,网络切片间的网络架构、网络功能是基本一致的,无法更好地适配不同场景的需求。因此,未来6G需要一个比5G更为完善的网络架构。

基金项目: 国家重点研发计划项目(2020YFB1806700、2022YFB2902104)

本文分别从3个不同的维度展望6G网络架构，包括总体系统架构、功能架构和组网架构，主要创新如下：

- 1) 提出总体系统架构设想，涉及三层、四面、云网运营和云网安全，满足新服务需求；
- 2) 提出至简网络功能架构，通过服务化和简化，降低网络复杂度；
- 3) 提出分层分布式组网架构，通过网络协同，提升网络性能以及灵活可扩展性。

1 6G网络总体系统架构

本节中，我们从系统框架的角度，分析6G网络总体系统架构。未来6G网络将是一个面向“连接+”设计的系统架构，具体如图1所示。

1) 三层

6G网络系统架构自下而上分别为云网资源层、网络功能层和应用使能层。

(1) 云网资源层：承载6G网络的基础设施和资源，资源类型包括各种计算、存储、网络、频谱以及各种专用设施。

(2) 网络功能层：6G网络核心的逻辑功能层，是实现6G网络核心功能的主体；在5G网络增强的控制面、用户面上基础上，还将引入数据面和智能面，以满足6G更为丰富功能需求。

(3) 应用使能层：聚合网络服务能力、通用的应用服务组件，通过能力开放、应用使能框架等方式为应用或周边生态提供服务，实现统一的应用使能管理。

2) 四面

面向“连接+”的功能需求，网络功能层的“四面”逻辑功能如下：

(1) 控制面：实现网络连接服务、智能服务、算力服务、感知服务等的统一控制功能。作为网络控制的中心，控制面将与其他层面密切协同，完成多接入融合控制、鉴权认证、移动性管理、会话管理、策略控制、人工智能（AI）任务调度、算力资源调配和管理功能等一体化的管控。

(2) 用户面：支持网络可编程，灵活定义数据处理策略，具体功能包括隧道管理、数据流识别、业务感知、确定性通信保

障、数据封装、数据转发及流量引导等功能。用户面主要完成用户会话数据的传递，在一些场景下，还可能实现环境物体感知数据、AI任务数据等各种数据处理及转发。

(3) 数据面：实现数据与业务逻辑进一步分离，减少数据和业务处理的紧耦合。在网络中引入单独的数据面，可以完成各种数据的统筹管理，并且数据面通过标准接口向控制面、用户面、智能面开放。各种数据中有静态数据以及动态的实时数据，例如用户签约数据、用户状态数据、网络状态数据、连接数据、资源利用数据等。

(4) 智能面：6G网络的智能中枢，支持核心网、接入网的全面智能。智能面既服务于6G网络自身的智能化，也服务于用户和业务应用的智能化。智能面提供网络AI相关的功能，包括数据建模、模型训练、推理决策、知识图谱、反馈与评估等。

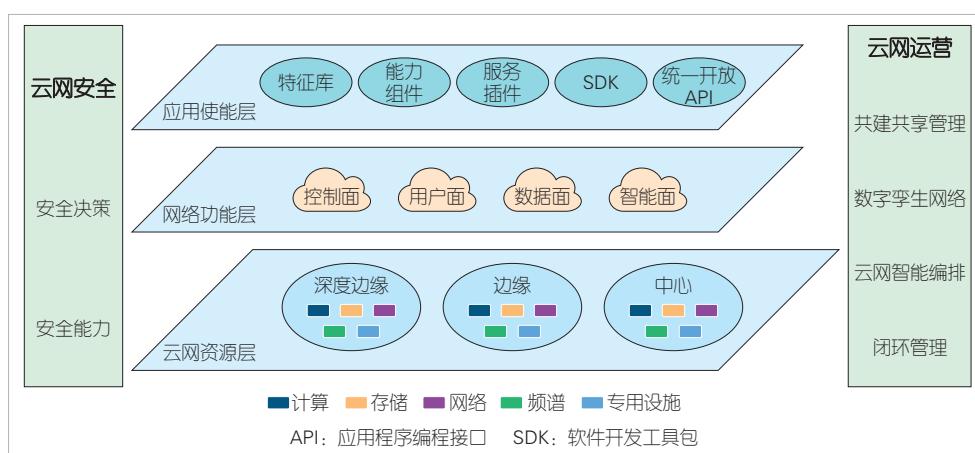
3) 云网运营管理

云网运营管理贯穿并覆盖网络的各个层面。在云网融合的演进趋势下，6G网络的运营管理除了拥有传统的网络管理、业务受理、计费结算等功能外，还将引入新的运营功能，具体如下：

(1) 云网智能编排调度功能：感知、识别和解析6G典型应用场景下多类型用户/客户的需求，按需生成网络策略，并利用编排器抽象资源状态，完成网络部署配置，实现端到端的服务自动化部署、负载均衡和服务保障、云边协同。

(2) 数字孪生网络管理功能：运用数字孪生技术构建数字化网络。网络与数字化网络之间进行实时数据交互，通过数字化网络对网络服务策略创建、下发、变更等进行有效的模拟验证，预先获得策略在物理网络上的实际执行结果。

(3) 基于区块链的共建共享管理：资源提供方和需求方共同组成一条链，资源提供方将资源信息记录在区块链上，资源需求方可通过区块链找到合适的资源，并且区块链所有



▲图1 6G网络总体系统架构

参与方都可通过共识机制对交易进行监督，实现公开透明、不可篡改和可追溯的资源动态共享交易。

4) 云网安全

云网安全贯穿和覆盖网络的各个层面。6G在延续传统移动网络的身份认证、密钥管理、空口安全等需求外，将针对新业务、新技术及安全形势的变化，提供动态自适应的安全能力、智能自主的安全决策、跨域协同的安全控制，具体如下：

(1) 安全能力：包括信任、监测和防御能力。信任能力是网络运作、业务正常开展的基础，也是风险识别、安全监测防御体系的基础；防御能力为网络提供安全保障，并通过监测能力动态、持续感知网络安全状态，为安全决策提供依据，当网络环境、安全状态发生变化时，做出相应指示。安全策略和控制指示安全防御做出策略调整，或调用安全防御能力进行响应、处置，直至恢复。

(2) 安全决策和控制：作为网络实例的安全能力决策、控制中枢和调度中心，支撑各安全能力面的运转和协作。安全决策既需要监测各能力的数据支持，同时也需要通过安全控制，将安全策略反馈给各安全能力，进而作用于网络。

2 6G网络功能架构

本节中，我们从网络功能的角度，分析6G网络功能架构。6G网络功能架构设计一方面需要在服务化架构（SBA）的基础上进行优化和增强，在提升网络服务的灵活性、弹性的前提下，力求通过网络简化和服务重构等增强服务效率、鲁棒性和可靠性。同时，将服务化的范围进一步向无线接入侧（例如无线网络和终端网络）延伸，实现移动领域端到端的服务化。另一方面，需要重点考虑如何满足未来新场景带来的大数据高效传输等需求。为此，未来6G网络功能架构将简化为网络控制、网络数据、网络用户面、网络智能、网络辅助5类服务，如图2所示。

1) 网络控制类服务

网络控制类服务的重点是服务及接口的重构。继承和优化5G网络的接入管理、移动性管理、会话管理、签约管理、鉴权、策略控制等基本连接服务，服务及接口基于功能、部署、开放等需求聚类或重构，解决架构复杂、互操作困难等问题；

增加接入适配服务，灵活兼容4G/5G、Wi-Fi、卫星、固网等多制式接入，提供通用、统一的服务能力；增加扩展服务，支持通信感知一体化（ISAC）、算网协同等新服务或定制服务的灵活、动态加载。

2) 网络数据类服务

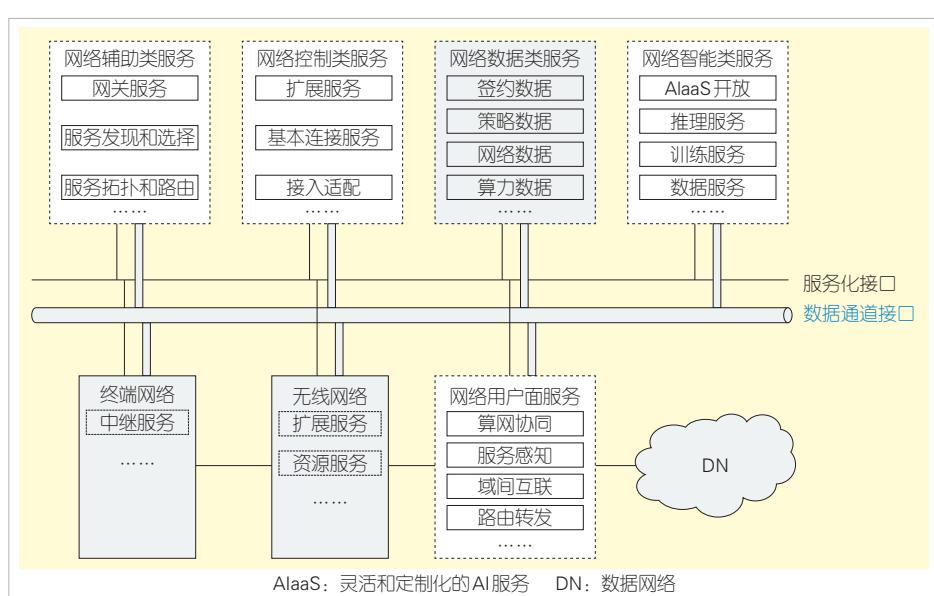
网络数据类服务聚焦数据与网络服务的进一步分离，配合增强的数据传输机制，形成数据面。网络数据类服务支持数据的逻辑统一、物理上可分布的管理和调用模式。除了传统的签约数据和策略数据外，该服务还新增网络数据、算力数据等基础数据用于其他服务的使用，实现通信和计算的融合；引入独立的数据通道，作为网络内部服务之间的另一个通用总线，与网络数据类服务协同，形成有机组织的数据面。数据通道支持网络内部服务与服务之间、终端与内部服务之间异构，以及大数据的存储、处理和传输等，满足ISAC、AI等新场景的数据需求。

3) 网络用户面服务

网络用户面服务重点增强分布自治、算网协同等带来的新功能。继承传统用户面对终端业务数据的路由和转发、策略执行等功能外，网络用户面服务还增加分布自治所需的自治域间路由、服务连续性需求，引入用户面可编程，实现用户面处理逻辑的灵活定义，同时支持并增强服务化接口，满足用户面的服务发现、控制、开放等。另外，网络用户服务还能增强服务感知和算网协同能力。

4) 网络智能类服务

网络智能类服务重点向分布式智能内生演进，已成为6G网络泛在的基础服务。网络智能类服务提供数据、训练、



▲图2 6G网络功能架构

推理等各类AI能力，对内将AI技术应用于网络，提升通信网络的智能化，对外开放灵活和定制化的AI服务（AIaaS）；通过独立设置+服务内嵌的混合方式，支持服务内部、自治域内、跨自治域等不同层级，跨终端、无线、核心网、业务、管理等多专业域的多节点协同，形成端到端、分布式的智能面。

5) 网络辅助类服务

网络辅助类服务提供分布、开放网络所需的支持。服务网关用于跨域能力开放等服务所需的映射和翻译；服务发现和选择用于自治域的域间和域内、网络域（核心网、无线网、应用等）之间的服务管理；网络拓扑和路由用于进行拓扑连接判断、路由配置等。

此外，在终端方面，随着近域通信等能力的成熟，终端的自组网能力及所需的策略控制和保障将得到加强，逐渐形成终端网络。未来，随着服务化技术的进一步演进，终端也可能按需支持服务化接口，例如终端对外提供中继服务，与核心网通过服务化接口连接。无线网络在集中式单元（CU）的控制面和用户面实现分离后，在满足服务指标的前提下，可在资源服务、AI和ISAC等新服务方面，逐步通过服务化接口对外开放。

6G网络功能架构采用双总线架构：一条是现有的服务化信令交互总线，延续5G已有服务化设计理念，将服务化的范围由核心网的控制面向用户面发展，同时进一步向无线接入侧延伸，例如无线网络和终端网络，从而实现移动领域端到端的服务化；另一条是新增的数据总线即数据通道，不同网络单元之间通过数据通道进行大数据传输，以达到数据的高效存取和处理的目标。服务化接口和数据通道可以配合使用，当网元之间通过一次信令交互可以完成所需数据交互时，使用服务化接口；当数据量过大或超出服务化接口处理能力时，通过服务化接口协商数据通道使用的具体协议，并根据传输数据的类型、数据量的大小、实时性要求等选择合适的传输协议。

3 6G网络组网架构

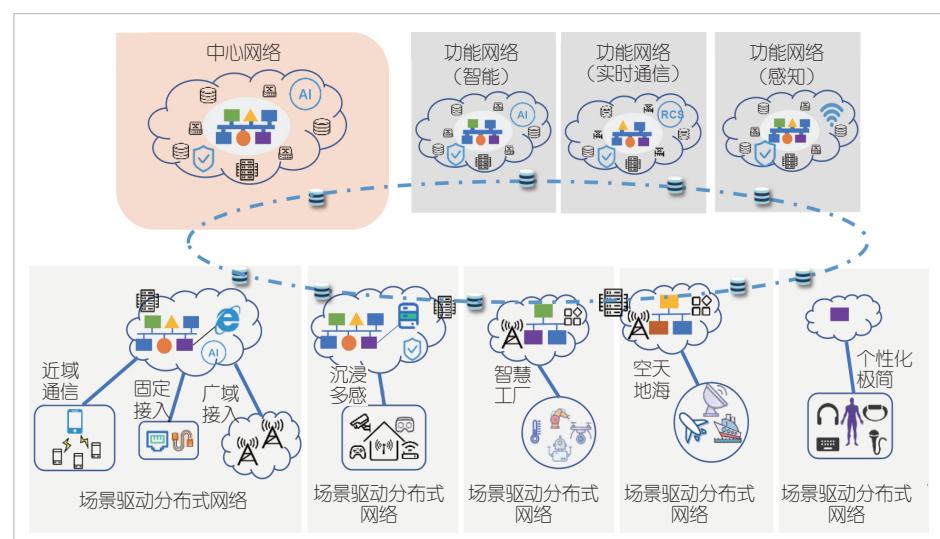
本节中，我们从组网的角度，分析6G网络组网架构。面向当前集中式架构带来的灵活扩展以及场景适配的挑战，未来6G网络将是一个分层分布式的组网架构，具体如图3所示。

6G网络部署架构引入了中心网络、分布式网络和功能网络，呈现分层分布式架构。其中，中心网络主要面向2C用户，满足广域覆盖需求；分布式网络主要满足各个场景下的基本业务需求，例如面向2C的低时延接入需求，以及面向2B的本地化接入服务；功能网络则满足用户的增值业务需求，包括实时通信网络、智能服务网络和感知服务网络等。

分布式网络支持面向不同应用场景、服务需求等，对网络进行场景化定制，灵活地选择网络接入方式、网络功能模块集，按需选择所需的计算资源、应用平台如移动边缘网络（MEC）等。此外，根据场景需求，分布式网络灵活地选择网络功能的部署位置，从而利用当前的分布式网络资源，向网络需求方提供多样化的、动态的网络服务。

功能网络可以根据需求选择部署方式，以实现网络的优化。为了进一步优化网络性能，满足不同需求，功能网络可以进行分布式下沉部署。中心网络与/或功能网络辅助分布式网络满足特定场景的服务需求。多个分布式网络可以共享一个中心网络或功能网络，例如，多个场景中的分布式网络可以共享同一个实时通信网络，感知服务网络或智能服务网络。这种共享可以减少冗余和重复部署，使得网络资源得到更有效的利用。

中心网络、分布式网络可以独立提供服务，也可以相互协同实现更完善的网络覆盖能力，以及无处不在、随时随地的互联网业务和定制化业务访问服务，还可以通过协同功能网络提供增值服务。例如，功能网络+2C下沉边缘网络可以共同为2C用户提供服务，功能网络+2B专网可以共同为2B



▲图3 6G网络部署架构

用户提供服务，具体如下：

1) 分布式网络独立自治

分布式网络具备独立运行的能力，能够独立完成网络内部的服务注册、发现与调用，独立提供服务；具备自治能力，能够实现网络功能及服务的自动化部署、即插即用，能够根据网络状态和应用需求调整网络资源，同时支持用户/客户对其进行管理、配置和监控。

2) 分布式网络与中心网络间协同

用户终端可以通过分布式网络接入到中心网络，访问原有的互联网业务。分布式网络用户可以从中心网络接入至自己所属的分布式网络中，访问定制化业务。此外，网间的互信授权还可以提供公网用户访问分布式网络业务、分布式网络用户访问公网服务的能力。

3) 分布式网络间协同

分布式网络的用户终端可以通过其他的分布式网络接入到自己所属的分布式网络，访问原有的分布式网络业务，也可以通过分布式网络间的互信授权实现分布式网络A用户接入分布式网络B网络，从而进一步提高网络的可靠性。

4) 分布式网络/中心网络与功能网络间协同

用户终端可以通过分布式网络或中心网络接入功能网络，访问功能网络提供的增值业务，解决用户终端对非基本业务的需求。例如，分布式网络或中心网络的用户A和用户B在通信时，可以通过功能网络提升通信体验。

5) 网络互联

6G网络需要支持网络间的发现、选择以及互联互通，且网络间的发现、选择与网元业务逻辑解耦。网络间支持相互认证、信令消息的访问控制验证，实现网络间的隔离和保护，具体可采用类似网络存储功能（NRF）的动态注册和发现机制、基于区块链/Service Mesh技术的分布式管理机制等。

4 结束语

ITU发布的《IMT面向2030及未来发展的框架和总体目标建议书》标志着6G研究重点正式从愿景需求转向技术方案落地。网络架构是6G研究的关键问题，是实现6G愿景的基础。

本文中，我们从网络系统、网络功能和组网部署3个方面阐述6G网络架构，包括增强网络逻辑功能、增强服务化、

简化网络功能和服务、设计分层分布式架构等。下一步，我们将研究连接、智能、数据等多维度协同调度，分布式网络协同，6G网络与5G网络的关系等6G网络架构及关键技术相关问题，携手学术界、产业界的伙伴共同推动6G技术进步和产业发展。

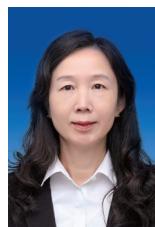
参考文献

- [1] ITU. Framework and overall objectives of the future development of IMT for 2030 and beyond [R]. 2023
- [2] 中国电信研究院. 6G愿景与技术白皮书 [R]. 2022
- [3] NGMN. 6G requirements and design considerations [R]. 2023
- [4] IMT-2030(6G)推进组. 6G典型场景和关键能力白皮书 [R]. 2022
- [5] IMT-2030(6G)推进组. 6G分布式网络技术的应用场景及需求研究报告 [R]. 2022
- [6] 中国电信, 中兴通讯. 未来移动核心网演进趋势白皮书 [R]. 2023
- [7] Next G Alliance. Next G alliance report: 6G distributed cloud and communications system [R]. 2022
- [8] Hexa-X. Final 6G architectural enablers and technological solutions [R]. 2023
- [9] Hexa-X. Hexa-X architecture for B5G/6G networks—final release [R]. 2023
- [10] 5G PPP Architecture Working Group. The 6G architecture landscape european perspective [R]. 2023

作者简介



刘玉芹，中国电信股份有限公司研究院6G研究高级工程师；主要研究领域为6G移动通信网络架构和关键技术，主要研究方向为6G分布式网络。



邢燕霞，中国电信股份有限公司研究院6G资深专家，高级工程师；长期从事移动核心网技术和业务研究工作，主要研究方向为6G网络架构；发表论文20余篇。



陈鹏，中国电信股份有限公司研究院6G研究中心主任，教授级高级工程师；长期从事无线通信研究和标准化相关工作，主要研究方向为下一代移动通信网络架构及关键技术；承担多项国家重大专项。

6G 网络架构和关键技术展望



Prospect of 6G Network Architecture and Key Technologies

王友祥/WANG Youxiang, 唐雄燕/TANG Xiongyan

(中国联合网络通信有限公司研究院, 中国 北京 100048)
(China Unicom Research Institute, Beijing 100048, China)

DOI: 10.12142/ZTETJ.202305005

网络出版地址: <http://kns.cnki.net/kcms/detail/34.1228.TN.20231016.1556.018.html>

网络出版日期: 2023-10-17

收稿日期: 2023-08-12

摘要: 在6G网络技术的发展进程中, 网络架构至关重要, 它是移动通信网络的基础和核心中枢, 决定了整个系统的效率和能力。分析了全球6G发展现状和5G-A网络架构演进, 提出了面向“新网络、新服务、新生态”的6G网络架构设计, 并介绍了6G网络发展的潜在关键技术, 包括空天地海一体、通感算一体、数字孪生、智能内生等。

关键词: 6G; 网络架构; 系统设计

Abstract: In the development process of 6G network technology, network architecture is very important, which is the foundation and core center of mobile communication network, and determines the efficiency and capability of the whole system. The global 6G development status and the evolution of 5G-A network architecture are analyzed, and the 6G network architecture design for "new network, new service, and new ecology" is proposed. Finally the potential key technologies of 6G network development are introduced, including air-space-ground integrated network, integrated sensing, communication and computation, digital twins and intelligent endogeneity.

Keywords: 6G; network architecture; system design

引用格式: 王友祥, 唐雄燕. 6G网络架构和关键技术展望 [J]. 中兴通讯技术, 2023, 29(5): 21–27. DOI: 10.12142/ZTETJ.202305005

Citation: WANG Y X, TANG X Y. Prospect of 6G network architecture and key technologies [J]. ZTE technology journal, 2023, 29(5): 21–27. DOI: 10.12142/ZTETJ.202305005

5G规模商用提升了用户通信能力和服务质量, 同时也促进了网络流量的快速增长。根据爱立信最新发表的移动报告^[1], 2022年第2季度至2023年第2季度, 移动网络数据流量增长了33%, 达到每月134EB。报告显示, 在智能手机使用量、移动宽带以及行业数字化的持续增长的共同作用下, 过去两年移动网络流量几乎翻了一番, 预计到2028年将增加到329EB。移动通信通常10年一个代际更替, 因此学术界和工业界开始探索下一代移动通信系统(6G)。人工智能(AI)、卫星互联网、云原生、算力网络^[2]等新技术的出现, 以及网络低碳绿色化的发展要求, 驱使通信网络朝着更强能力、更高能效的方向发展, 以更好地满足不断产生的各类业务的服务要求。

目前各国正在加紧6G预研。从2020年到2022年, 欧美日韩等国家和地区纷纷成立了6G研发行业联盟, 包括美国NextG联盟、欧盟6G-IA等^[3]。2022年2月, NextG联盟发布了《6G路线图》, 确立了6G远景目标、技术方向和发展路线。

欧盟2021年启动Hexa-X 6G研究项目, 聚焦6G的应用场景与关键技术; 2023年1月启动第2阶段研究项目Hexa-X-II, 聚焦6G的系统化和预标准化研究工作。日韩也专门设立研究基金, 推动6G核心技术研究。中国在2019年由工业和信息化部牵头成立了中国IMT-2030(6G)推进组, 聚焦中国6G关键技术研究, 陆续发布了《6G典型场景和关键能力白皮书》《6G总体愿景与潜在关键技术白皮书》《6G前沿关键技术研究报告》《6G网络架构愿景与关键技术展望白皮书》《6G无线网络架构和功能技术研究报告》等近20份相关技术白皮书。2022年国务院出台了《“十四五”数字经济发展规划》, 提出前瞻布局6G网络技术储备, 加大6G技术研发支持力度, 积极参与推动6G国际标准化工作。

2023年6月, 国际电信联盟(ITU)完成了《IMT(International Mobile Telecommunication)面向2030及未来发展的框架和总体目标建议书》。未来6G将包含沉浸式通信、超大规模连接、极高可靠低时延、通信与智能融合、感知与通信融合、泛在连接六大典型应用场景。通过与其他技术的结合, 如先进计算、AI、大数据、区块链等, 6G将实现网络

基金项目: 国家重点研发计划项目(2022YFB2902103)

性能和服务能力跃升，全面引领经济社会数字化智能化绿色化转型^[4]。该建议书明确了6G的15个能力指标，包括9个6G增强能力和6个6G的新能力。与5G相比，IMT-2030能力指标分为两类，即针对IMT-2020增强的功能和支持IMT-2030扩展使用场景的新功能。每种功能在不同的使用场景中可能具有不同的相关性和适用性。其中，针对IMT-2020增强的功能包括：峰值速率、用户体验速率、频谱效率、区域流量密度、连接数密度、移动性、时延、可靠性、安全隐私韧性性能，支持IMT-2030扩展使用场景的新功能包括：覆盖、感知相关指标、AI相关指标、可持续性性能指标、互操作、定位。

基于全新的能力体系，6G将摆脱以往通信网络“纯连接”的特性，打造“联接+计算+智能”的通信新范式，成为连接真实物理世界与虚拟数字世界的纽带，实现人机物智能互联、协同共生，以及实体经济和数字经济的全面融合，推动社会普惠智能、绿色健康可持续发展，满足经济社会高质量发展需求。

1 5G网络架构和演进

相比于4G，5G采用了全新的网络架构，核心网引入了服务化（SBA）设计，无线接入网采用集中单元（CU）和分布单元（DU）分离的两层架构设计。5G核心网采用虚拟化方式实现。控制面与用户面完全分离，用户面可以基于业务需求灵活部署，控制面采用SBA架构，功能以服务方式定义，相对独立，采用轻量高效的服务调用接口（即服务间通信接口），以应用程序编程接口（API）方式对外呈现。为了满足业务多样化的需求，5G定义了多接入/移动边缘计算（MEC）技术，使得用户可以就近访问业务；定义了网络切片技术，可以使用一套物理网络服务于多个不同的应用场景。5G新空口（NR）以分组数据汇聚协议（PDCP）/无线链路控制（RLC）层为界，将基站分为CU和DU两个功能实体。其中，CU承担无线资源控制（RRC）/PDCP层功能，DU承担RLC/媒体接入控制（MAC）/物理层（PHY）功能，分离架构为运营商提供了多种部署形态选择。

5G-Advanced是基于5G的演进和增强，具体包括：（1）切片能力增强，可基于统一的网络基础资源及设施，隔离出多个虚拟的端到端网络，并保证各虚拟网络的业务隔离性；（2）网络数据分析功能（NWDNF）增强，聚焦于面向商用部署的分布式智能架构、新型分析机制和应用案例的研究，并拓展与周边系统之间的智能协同；（3）天地融合网络架构增强，包括覆盖增强、移动性增强、10 GHz以上甚小口径卫星终端站（VSAT）宽带数据业务等；（4）面向垂直行业

的5G局域网（LAN）、非公共网络（NPN）、确定性网络、高精度定位等能力增强；（5）引入了下一代实时通信（NG-RTC）能力，对IP多媒体系统（IMS）网络架构及功能进行增强，以满足实时通信新业务特性要求；（6）在无线侧重点是对无线空口的承载能力和智能化的增强，包括Sidelink增强、广播多播增强、针对扩展现实（XR）的增强、非地面网络（NTN）增强、5G RedCap、上行链路增强等^[5-6]。

5G-Advanced是从5G演进到6G的关键阶段，具有“承上启下”的作用。6G网络架构设计要充分吸收5G及5G-Advanced网络的实践经验，继承5G及5G-Advanced成熟技术和理念，在最大化吸收网络演进新技术能力、适配新业务需求的同时，兼顾节能增效、降低成本，并保持与现有网络的兼容、继承和协同。

2 6G网络架构设计

网络架构是移动通信网络的基础和核心中枢，决定了整个系统的效率和能力。未来，6G网络一方面将向边缘网络空间和空天地海不断延伸，满足对天基、空基、地基等多种接入方式，固定、移动、卫星等多种连接类型的接入需求，实现空、天、地、海一体化无缝覆盖，向全域万物智联的方向迈进；另一方面将实现网络能力拓展，即从单一的通信连接能力，拓展到通信、感知、计算、数据、智能、安全等多维的能力。

2.1 架构设计理念

2021年，中国联通发布CUBE-Net 3.0网络创新体系^[7]，提出了构建面向数字经济新需求、增强网络内生能力、实现“联接+计算+智能”融合服务的新一代数字信息基础设施。这是一个全新的网络架构设计理念。基于此理念以及未来6G时代的需求和愿景，6G网络架构设计有以下三大理念^[8]：

1) 面向“超越连接”的设计

网络提供的服务将从单一的连接服务，扩展到智能、感知、数据、计算、安全新服务。相关设计理念体现在以下方面：

内生：现有网络架构，以大带宽、低时延的连接功能为主，其他功能的引入和实现主要通过外挂式的方式完成，无法满足6G多样的业务需求和动态复杂环境需求。因此，6G网络架构需要注入多种功能内生和融合基因，从网络功能、接口、协议栈等方面端到端地支持智能、算力、安全等内生能力。

协同：现有面向连接的网络架构，提供了基于会话的管控和连接服务质量（QoS）保障机制。当6G网络需要超越连

接提供各类新业务时，需要提供超越会话的协同机制，在网络架构层面引入新的管控量纲，基于任务协同新业务所需要的连接、数据、计算等多个维度资源，以此实现新业务的管控和QoS保障。

分布：6G多样化新业务所需的连接、计算、数据等能力来源于不同区域、不同资源池、不同组织，6G网络需要调度和联接上述多种不同的网络能力，提供用户就近、随用户移动的、分布式的网络服务，满足用户对6G融合业务的诉求。

2) 面向“多元用户”的设计

网络从为终端用户提供服务转变为向行业用户、OTT(Over The Top)应用、子网络等泛在用户提供综合服务。相关设计理念体现在以下方面：

多元：随着新业务的引入6G网络的服务对象将得以扩充，用户不仅是传统使用连接服务的终端用户，还包含使用网络计算服务、数据服务、感知服务等新业务的多元用户。

弹性：6G移动网络为了向多元用户提供差异化的网络服务及就近随行的网络能力，需要提供弹性的网络功能架构、简捷的用户网络提供方式，降低网络部署运维成本。

定制：6G网络丰富了网络服务的内涵，可以为多元用户提供不同的服务类型与QoS保障，需要允许用户高效灵活地定制专属网络，提供差异化的用户网络服务，满足不同个性化的网络需求。

3) 面向“平台即服务”的设计

网络从提供单边的连接管道服务变成面向多元用户提供多边服务的使能平台。相关设计理念体现在以下方面：

开放：网络的参与者（终端用户、行业用户和合作伙伴等）基于统一的服务使能平台开放自身支持的能力/服务，同时，可根据领域知识将不同的原子能力组合/抽象成新的API，并开放给用户，以降低网络定制成本，赋能多样化生态。

可编程：基于多样化业务，对控制面功能、用户面功能、计算功能、数据功能等功能进行灵活编排，实现网络能力、部署方式、执行路径等可编程，满足差异化的业务QoS需求。

多边：多元用户既可以是网络功能/服务的消费者，也可以是功能/服务的生产者。用户与网络建立灵活友好的交互方式，实现用户与用户、用户与网络之间的多边能力资源的协同。

2.2 网络架构设计

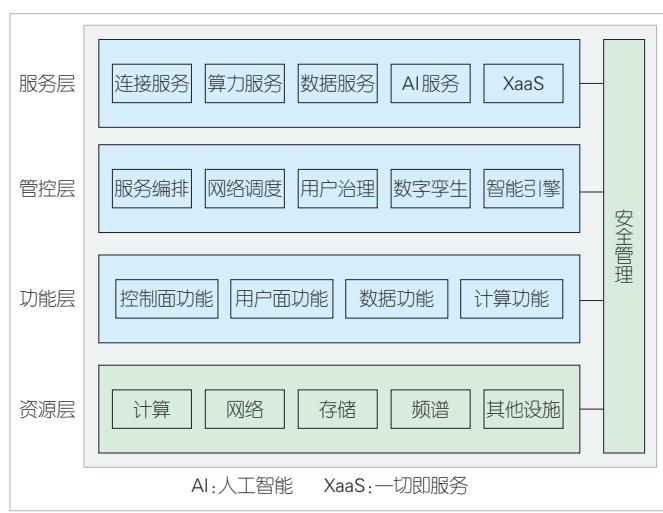
基于前述的架构设计理念，我们提出了面向“新网络、

新服务、新生态”的层次化的数智服务使能平台架构。如图1所示，该架构上下共分为4层，自下而上分别为资源层、功能层、管控层和服务层^[8]。核心功能如下：

1) 资源层：为6G网络功能部署、运行和服务提供支撑，主要包括计算、存储、网络和频谱等各种基础设施，是整个网络运行的基础。资源层物理设施实体分布在空天地海、云边端、广域局域等场景空间中，除了运营商自建设备外，还包括友商共建共享资源，或者产业上下游合作伙伴/多元用户等不同组织归属的资源，为6G网络的泛在接入、普惠开放能力提供保障。

2) 功能层：是6G网络的执行层，对下，根据需求对资源层物理设施进行多维资源的互联、组织、协同、调度构建具备不同网络能力的业务逻辑；对上，为管控层提供数据、计算和决策执行能力支撑。按照业务类型，功能层分为控制面功能、用户面功能、数据功能和计算功能，根据管控层的智能决策拉通物理资源，为网络和服务层提供接入控制、连接传输、数据采存、算力纳管互联和计算执行等业务服务能力支持。

3) 管控层：是6G网络的大脑，接受服务层的请求，在对服务请求进行多维度智能分析的基础上，为服务和应用作出智能决策，指导功能层操作执行，为多元用户提供开放、灵活、多边的网络服务能力，包含服务编排、网络调度、用户治理、数字孪生、智能引擎等。服务编排可将服务层的需求快速解析为单项服务能力和服务质量要求；网络调度快速按需创建网络功能，创建网络路由等，动态调整网络各节点的算力、带宽、存储等资源，对网络进行弹性伸缩处理等；用户治理用于多元用户注册、生命周期管理、能力资源管理协同度量等；数字孪生为现实网络提供虚拟的孪生映射，为



▲图1 6G网络系统架构图

网络提供分析、仿真、诊断、预测、优化等仿真模拟和可视化支持；智能引擎为服务编排、网络调度、用户治理等功能提供意图分析、运行数据分析、运营数据分析等智能分析能力，与数字孪生联合模拟仿真，促进网络的自智等级提升。

4) 服务层：对下层的网络功能进行提取、封装和组合，为内部业务或第三方应用按需提供能力或服务。服务层可开放运营商网络的连接服务、算力服务、数据服务、AI服务、定制化服务等能力，为多样化业务场景下的多元用户，带来更加深入的算网业融合能力和更好的业务体验^[9]。

6G系统架构中的功能层包含4个基本功能面，分别为控制面功能、用户面功能、数据功能和计算功能。这些功能面分别负责网络控制、路由转发、数据管理等网络功能，以及数据、智能等内生功能。功能层作为6G网络架构中的执行层，包含的4个功能面既是功能层内提供控制、用户、数据和计算4类业务逻辑的功能集合，也是6G网络对外提供接入、连接、数据、计算4类端到端服务能力的体现。

- 控制面功能：支持空天地海泛在接入控制能力，以及对这些异构接入间的融合调度协同能力；在传统连接服务的基础上，进一步提供计算、数据、感知等服务的一体控制与移动性管理；支持柔性编排、按需下沉，实现分层分布式部署。此外，控制面功能可以为用户提供更灵活的网络协同能力。

- 用户面功能：支持功能集模块化和可编程，增加业务感知、内生算力、智能分析、确定性转发、安全管理等能力；通过灵活的协议栈编排配置，实现通感算智数安等多维业务的高性能分级数据转发和处理；支持跨域协同，与承载网对接，保障确定性的端到端QoS；通过扩展用户面协议栈或调用AI能力，使能用户面功能感知业务信息，实现智能业务感知与差异化QoS保障；支持泛在互联，为新业务提供更多维度新业务数据的转发与传输，连通周边的新业务网络功能。

- 数据功能：具备数据采集能力、数据处理能力、数据存储能力和数据协同能力的6G数据功能将面向多元的功能或业务实体，在保障数据隐私、安全及可靠性的同时，实现对数据的精细化采集、高效传输、分布式协同和弹性存储，统管数据的全生命周期，从而提升海量网络数据（如内部数据、网络数据、感知数据和计算数据等）的流动效率，保障用户服务质量，增加数据服务的安全性。

- 计算功能：6G网络功能将向对外提供算力服务的新形态网络功能方向发展，结合人工智能技术实现6G网络智能内生。通过计算与通信网络的深度融合，对外提供计算服务。6G网络计算功能包含计算控制、计算执行、计算互联

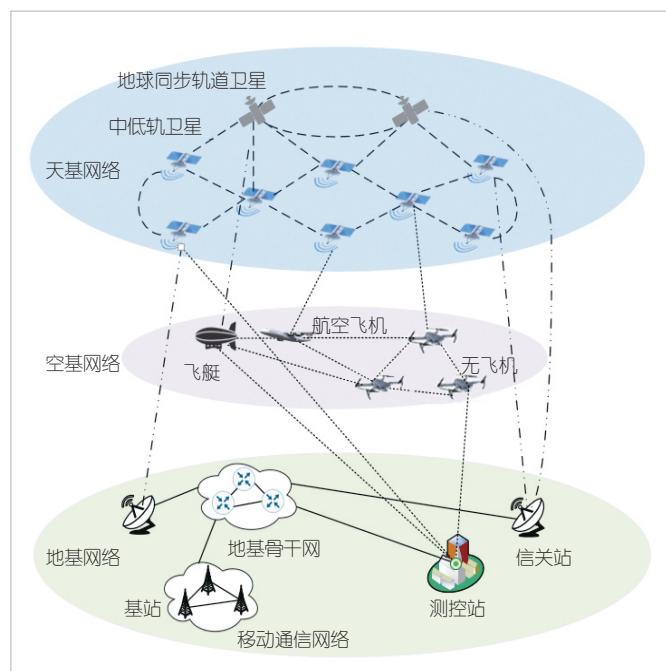
等功能，以满足未来多种应用场景的算力服务需求。同时，为了更好地实现网络的智能化与可编程化，使网络可以适应不同应用场景，计算能力服务化以及计算功能的泛在化部署是6G网络所必需的网络能力。

此外该架构还包含一个贯穿各层的、内生安全管理功能。安全是发展的基础，6G网络的安全设计也是6G网络架构的重要组成部分。在本架构中，安全管理将贯穿于资源层、功能层、管控层、服务层的全层次体系，自底向上、端到端构建内生安全能力，为6G网络提供全生命周期的安全防护保障。

3 6G网络关键技术

3.1 空天地一体化网络技术

“全球覆盖、随遇接入、按需服务、安全可信”的空天地一体化网络是未来6G重要的演进方向，其愿景是：构建以地基网络为基础、天基网络（含通导遥卫星）和空基网络为补充和延伸的立体网络结构，形成跨地域、空域、海域的天地一体融合通信网络，满足广域智慧连接和全球泛在无缝接入需求，实现统一高效的资源调度与网络管控，进而为广域立体空间范围内的应用提供全球时空连续通信、高可靠安全通信、区域大容量通信、高机动全程信息传输等能力。这对提高应急水平、维护国家安全、保障国计民生、促进经济发展具有重大意义^[10-11]。面向6G的空天地一体网络架构如图2所示。



▲图2 6G空天地一体网络架构示意图

面向6G的空天地一体网络是天基多层子网和地面蜂窝多层子网等多个异构网络，在体制、协议、网络、业务、终端等方面的深度融合，需聚焦解决融合网络多层次立体、动态时变等关键问题，构建空天地一体网络控制能力，实现网络架构、空口传输、频率管理、星地组网等关键技术突破及分阶段部署。6G空天地一体网络主要包括：

统一网络架构：通过地面网络和非地面网络在系统架构、技术体制、接口协议等层面进行多域多维度柔性融合，实现空天地网络的高效协同，提升端到端管理效率。

统一空口协议：形成卫星与地面蜂窝通信的统一空口协议，支持多种业务传输。

频率协调管理：通过卫星与地面频谱协调管理技术，解决星地网络同频或邻频部署时网络共存、网络协同等方面的问题，提高频谱利用率。

星地融合组网：通过空天地一体网络新型组网技术研究，解决业务连续性和一致性、移动性管理、QoS保障等关键问题，提升空天地业务服务能力。

3.2 分布式网络技术

ITU定义的6G六大场景包括沉浸式通信、超大规模连接、极高可靠低时延、人工智能与通信融合、感知与通信融合和泛在连接。这意味着未来网络承载的用户和业务种类、协议数量、网络功能以及各功能间的连接数量等都将成倍增长。如果仍沿用目前集中式的架构设计，网络将变得越来越复杂。因此，为追求更极致的网络性能、更好的业务体验，分布式网络将成为6G网络架构演进的重要趋势之一，能够根据未来多样化的业务和需求，为用户提供不同位置按需部署、灵活组织编排、弹性可扩展的分布式网络部署服务能力，和跨网互通协同能力^[12]。6G分布式网络涉及以下关键技术：

广泛分布的多样接入模式融合：6G网络由大量用户网络（子网）构成，不同子网的用户可能在不同位置并以各种不同的方式接入到6G网络。因此，分布式网络具有融合多样化的接入模式能力，能够协调用户多接入路径，向用户提供更加可靠、高效的连接通道。

连接功能的按需分布式部署：由于不同用户对连接能力的差异化要求，分布式网络需要在相对广域的范围内灵活动态部署控制面和用户面，为用户提供最优的连接能力。

计算功能的分布式多方协同：6G网络将形成超级的网络计算服务，计算功能逐步下沉至网络边缘为用户所用已成为普遍的趋势，实现对分布式计算（包括算力、算法和数据等）、分布式网络和分布式业务的统一高效一体化编排与动态调度，满足定制化的网络需求，提供面向用户及业务的内

生计算能力。

数据功能的按需编排：数据作为未来网络的关键要素之一，将呈现显著的分布式特征，这表现在数据源、数据消费者、数据处理节点方面。6G网络需要提供具备处理分布式数据能力的数据服务框架，支持对数据功能的按需编排，高效应对复杂的数据应用模式，满足不同用户的数据处理需求。

基于共识机制的去中心化信任：6G分布式网络将构建一种去中心化的信任架构，保障多方信任快速准确地建立，为构建和维护稳定、公平的6G开放合作的商业生态提供有力的安全支撑。

因此，未来网络的场景需要兼顾覆盖和容量，将形成一种全新的分布式网络部署形态，其研究核心是：如何设计场景定制化的柔性至简子网方案，各分布式子网间如何实现互联互通和业务连续性保障，用户如何快速便捷获取和定制子网服务、以及跨运营商的共享和协同。业界需要从组网模式、网络联邦管理技术、网络开放接口和跨运营商网络接口等多方面进行增强研究。

3.3 内生AI技术

AI是6G的核心技术之一，AI与移动通信网络的深入融合一方面推动6G网络向自智化方向演进，提升运营商的生产和运营效率；另一方面支撑6G网络以AI即服务（AIaaS）的形式提供能力，将网络中的AI能力作为服务开放给各类行业和普通消费者，满足其对AI的功能、性能、隐私和个性化的需求。在面向6G的智能化演进过程中，网络将通过增强分布式AI框架、以任务为中心的AI能力，以及意图网络等，实现网络内生的智能能力，提供更灵活、高效、泛在的6G智能化解决方案。

分布式AI：基于NWDAF的5G智能化解决方案仍属于集中化、外挂式方案，在数据处理时延、数据传输消耗方面存在进一步提升和优化的空间。面向6G时代大规模、高性能、低时延的AI业务需求，AI能力的下沉成为网络演进的必然趋势。因此，6G网络应支持分布式AI能力部署，通过算力、模型等智能要素的下沉，实现数据的就近采集和就近处理，提升实时智能服务能力；同时，通过算法、模型、数据的分布式实现，进一步结合联邦学习、多任务并行及协同处理等能力，完成模型协同训练、数据协同处理等复杂AI任务的分解、处理和聚合，解决集中式AI所存在的成本、能耗、效率等方面不足，实现网络智能化资源高效利用。

任务为中心的AI：传统通信系统以通信连接为中心进行设计，主要目的是为数据传输提供连接并保证业务体验。

而AI类的处理过程则更加复杂，整体过程除了连接以外，还涉及计算、数据、算法等智能要素。为使6G网络具备内生智能能力，6G网络需引入新的资源维度并设计相应的管控机制。此外，为了AI业务，网络架构的设计思路应需要从以连接为中心转变为以任务为中心。其中，任务是指网络智能能力所涉及的多节点场景下连接、计算、数据和算法资源的协同和调配，以共同完成某个特定的目标，如AI推理、AI训练、计算等任务类型。

在以任务为中心的思路下，6G网络通过对智能管控框架的设计和编排，实现智能要素协同、多层次多节点协同，以任务的粒度完成智能业务处理，并进行任务级的QoS保障，实现6G普惠及内生智能需求。

3.4 通感算一体技术

利用软硬件资源的协同与共享，通信、感知和计算融合一体可以实现多维感知、协作通信、智能计算功能的深度融合和互惠增强，从而实现对物理世界的观测和采样，打开物理世界与数字世界融合的通道，提供定位、测距、测速、成像、检测、识别等多元化能力，极大满足超高分辨率和精度的应用需求。

未来6G系统的感知和计算功能，不仅可以提供对外的业务服务，同时也可以服务于6G网络本身。感知、计算（包含AI）和通信三者之间可以相互辅助、功能增强。6G系统通过感知服务对物理世界进行采样，构建数字世界，并通过通信服务为物理世界提供连接，再通过计算服务对通信和感知的过程及数据进行处理。如果以数据为中心，那么感知就是数据采集，通信则是数据传输，计算是数据处理。因此，除了信息传递者之外，6G还将扮演信息生产者和信息加工者的角色。

通感算一体化网络架构是集通信、感知、计算和应用为一体的系统架构。无线架构设计中，无线空口波形设计是一体化技术的核心之一，其目标是基于同一套射频收发设备和相同频谱，同时实现无线电通信和雷达感知的功能。这主要包括三大技术路线：基于通信波形的一体化波形、基于感知波形的一体化波形以及基于全新的通感融合的一体化波形。此外，从组网方面考虑，针对不同业务需求可以采用多频段协同感知，实现覆盖能力、感知精度的提升；通过组网多节点间协同感知，解决单个节点感知的覆盖能力和精度的局限问题，提升感知性能；通过多模式感知方式的协同，构建全方位的感知全域框架，实现对周边环境的精细精确感知^[13]。

通感算一体化架构需要根据业务考虑网络资源和算力资源的分配。管控层根据服务层的业务需求和指标映射，进行

业务编排，按需分配资源，形成网络配置要求，实现网络灵活互联和功能交互协作。基于网络功能层的逻辑分工，形成跨层的通感算融合设计和统一编排管理，并结合算力、网络等资源的实时状态合理分配计算任务，形成数据处理一体化、资源分配一体化和服务实现一体化的互惠共生架构^[14-15]。

3.5 数字孪生技术

数字孪生的概念随着数字化的演进被提出，是数字化的高阶形态和必经之路。数字孪生通过数字模型和实时数据构建现实实体的虚拟镜像，可以对物理世界的物理对象、过程或系统进行仿真和预测。数字孪生将物理世界和数字世界相结合，实现对物理世界全生命周期的虚拟验证、决策支持和优化控制。

随着移动通信技术的发展，6G网络作为下一代移动通信标准，将为人们带来更快速、更可靠、就近随行的极致体验。然而，6G网络的实现面临一系列挑战，如多元用户的泛在接入、深度协同、智能内生等需求促使网络规模进一步提高，融合接入、传输、计算、数据等服务能力的大规模网络的复杂性是一个关键问题。为了应对这些挑战并推动网络的进一步发展，引入数字孪生技术是一种有效解决方案。

6G网络通过数据采集、数据建模、仿真推演、智能决策、智能管控等关键控制环节构建6G数字孪生网络，赋能网络自智发展。数据采集是构建数字孪生的基础，孪生网络需要具备全景多维度数据采集能力，实现对物理世界网络的实时数据（状态）和非实时数据（状态）的感知；模型是构建数字孪生的核心，基于感知的网络数据，将网络从物理实体向虚拟空间映射，构建与物理实体一致的孪生数字网络；数字孪生通过轻量化的、灵活的、能真实刻画网元/网络运行的模型，再结合实时的真实网络数据，通过动态配置的规则组合成具体的仿真流程，实现快速仿真推演网络未来的变化，以达到评估和预测等目的；智能决策利用AI和知识图谱等技术对仿真流程和孪生网络模型进行分析，判断和识别出问题的类型及所在位置，并构建价值函数实现方案寻优和决策推荐，同时通过分析实时的网络数据和状态信息，并结合知识图谱识别出潜在的问题，预测网络的未来发展趋势；最后，基于决策算法对方案进行评估、优化和推荐。智能管控对智能决策形成的方案进行调度执行，通过指令通道完成解决方案的自动下发执行，实现网络参数自动配置，支持业务自动开通、自动故障识别和恢复、自动服务等级协议（SLA）保障等自智网络业务。智能管控技术结合智能决策技术，通过指令通道对物理网络进行实时或非实时的管理控

制，实现以虚控实。在将孪生决策层的指令下发到真实物理网络后，物理网络通过数据反馈给孪生网络进行策略效果评估，并再次触发决策分析优化，实现虚实闭环，最终实现高阶自智网络。

4 结束语

网络架构是6G能否显著提升网络能力、赋能社会发展的核心。6G在传统连接能力基础上衍生的感知、智能、计算、安全等多维能力，对网络架构提出了颠覆传统架构的创新需求。目前，空天地海一体、通感算一体、数字孪生、智能内生等关键技术方向已逐渐形成共识，但是如何将这些潜在技术有机融合形成6G网络的总体架构设计还在探索阶段。这需要产业界携手共同努力，形成技术共识，促成网络架构达成统一，为6G发展奠定基础。

参考文献

- [1] 爱立信. 爱立信移动市场报告 [R]. 2023
- [2] 张宏科, 权伟, 刘康. 算力网络研究与探索 [J]. 中兴通讯技术, 2023, 29(1): 1–5. DOI: 10.12142/ZTETJ.202301001
- [3] 王亦菲, 闻立群, 李明豫. 全球6G产业及政策进展研究 [J]. 信息通信技术与政策, 2022, 48(9): 71–75
- [4] ITU. Framework and overall objectives of the future development of IMT for 2030 and beyond [EB/OL]. (2023-06-28)[2023-09-15] <https://www.itu.int/md/R19-SG05-C-0131>
- [5] 秦鹏太, 李爱华, 姜怡, 等. 5G-Advanced核心网技术综述 [J]. 移动通信, 2022, 46(1): 58–66
- [6] 夏旭, 齐文, 王恒, 等. 5G-Advanced网络及服务演进需求探讨 [J]. 移动通信, 2022, 46(1): 15–19
- [7] 中国联通. 中国联通CUBE-Net3.0网络创新体系白皮书 [R]. 2021
- [8] 中国联通. 6G中国联通6G网络体系架构白皮书 [R]. 2023
- [9] 唐雄燕, 张帅, 曹畅. 夯实云网融合, 迈向算网一体 [J]. 中兴通讯技术, 2021, 27(3): 42–46. DOI: 10.12142/ZTETJ.202103009
- [10] IMT-2030(6G 推进组). 6G网络架构愿景与关键技术展望白皮书 [R]. 2022
- [11] IMT-2030(6G 推进组). 6G总体愿景与潜在关键技术白皮书 [R]. 2021
- [12] IMT-2030(6G 推进组). 6G 分布式网络技术的应用场景及需求研究报告 [R]. 2022
- [13] IMT-2030(6G 推进组). 通信感知一体化技术研究报告 [R]. 2022
- [14] 中国联通. 中国联通算力网络架构与技术体系白皮书 [R]. 2020
- [15] 中国联通. 中国联通6G通感算一体化系统架构与关键技术白皮书 [R]. 2023

作者简介



王友祥, 中国联通研究院高级工程师;主要从事移动通信网络架构和关键技术研究和部署应用等工作。



唐雄燕, 中国联通研究院副院长、首席科学家, “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, 工业和信息化部通信科技委委员兼传送与接入专家咨询组副组长, 北京通信学会副理事长, 中国通信学会理事兼信息通信网络技术委员会副主任, 中国光学工程学会常务理事兼光通信与信息网络专家委员会主任, 国际开放网络基金会ONF董事;拥有20余年电信新技术新业务研发与技术管理经验, 主要专业领域为宽带通信、光纤传输、互联网/物联网、SDN/NFV与新一代网络等。



6G 网络架构研究进展及建议

6G Network Architecture: A Survey

谢峰/XIE Feng^{1,2}

(1. 中兴通讯股份有限公司, 中国 深圳 518057;
2. 移动网络和移动多媒体技术国家重点实验室, 中国 深圳 518055)
(1. ZTE Corporation, Shenzhen 518057, China;
2. State Key Laboratory of Mobile Network and Mobile Multimedia, Shenzhen 518055, China)

DOI: 10.12142/ZTETJ.202305006

网络出版地址: <http://kns.cnki.net/kcms/detail/34.1228.TN.20231020.1241.002.html>

网络出版日期: 2023-10-20

收稿日期: 2023-08-12

摘要: 6G 网络架构的研究包括组网生态、面向高性能和能效的软硬件联合设计通信架构、云原生/算力网络和服务架构、新能力（可信安全内生和智能内生）架构 4 个维度。这 4 个维度与 ITU-R IMT-2030 (6G) 框架文件中的场景和指导原则密切相关。对中国、欧盟、美国三方的主要研究进展分别做了介绍和对比分析，并给出了一些关于中国 6G 网络架构研究的建议。

关键词: 6G 网络架构；生态；性能和能效；云原生；算力网络；服务架构；可信安全内生；智能内生

Abstract: The research on 6G network architecture can be classified into four perspectives: the networking ecology, high-performance and energy-efficient software-hardware co-design communication architecture, cloud-native/computing network and service architecture, and new capability architecture (trust&security-native, AI-native). These four perspectives are closely related to the use cases and overarching principles of the ITU-R IMT-2030 (6G) framework. A compact but comprehensive survey on the progress of research on 6G network architecture by China, Europe Union, and USA is introduced, along with some comparison and analysis. Finally, some advice for the research on China's 6G network architecture is provided.

Keywords: 6G network architecture; ecology; performance and energy-efficiency; cloud-native; computing network; service architecture; trust&security-native; AI-native

引用格式: 谢峰. 6G 网络架构研究进展及建议 [J]. 中兴通讯技术, 2023, 29(5): 28–37. DOI: 10.12142/ZTETJ.202305006

Citation: XIE F. 6G network architecture: a survey [J]. ZTE technology journal, 2023, 29(5): 28–37. DOI: 10.12142/ZTETJ.202305006

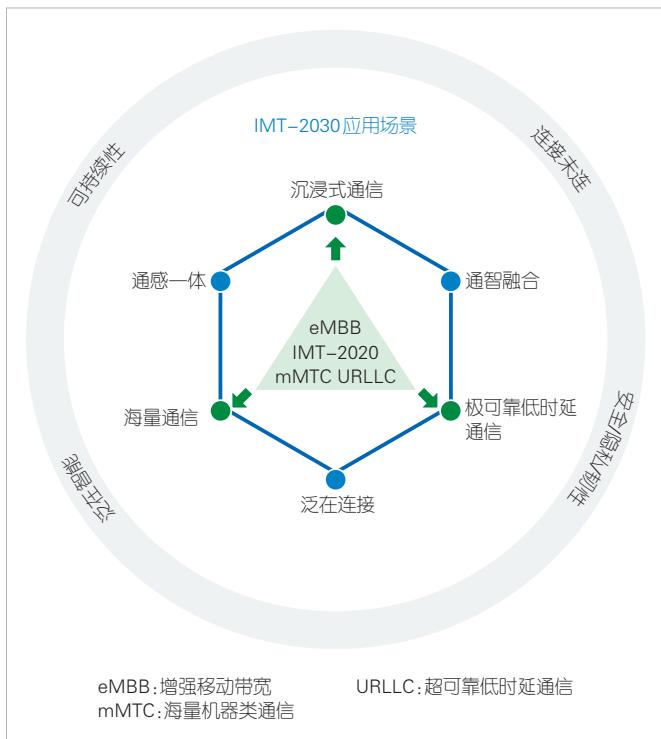
1 6G 需求和愿景

国际电信联盟无线电通信部门 (ITU-R) 在 2023 年 6 月正式发布 6G 框架文件^[1]。该文件涉及 6G 场景和关键能力，基本上涵盖了全球各主要国家和主要厂家的观点。如图 1 所示，6G 场景包括 5G 三大场景的扩展：从增强移动宽带 (eMBB) 扩展到沉浸式通信，从超可靠低时延通信 (URLLC) 扩展到极可靠低时延通信 (HRLLC)，从海量机器类通信 (mMTC) 扩展到海量通信。此外，6G 场景还包括通感融合、通智融合，以及面向尚未被覆盖或很少覆盖地区的泛在连接。同时，国际电信联盟 (ITU) 还给出了四大指导性原则：可持续性、连接未连（泛在连接场景与此呼应）、安全/隐私/韧性、泛在智能（通智融合场景

与此呼应）。

三大扩展场景涉及对先进性能指标（例如速率、时延、可靠性、连接密度等）及其灵活组合的要求。这对空口、设备软硬件以及网络架构提出了要求：用户面软硬件联合设计、可编排可编程网络、控制与转发分离（控制面/用户面分离）是其支撑技术。新的三大场景中，通感融合不仅为运营商网络、行业网络和终端提供了新的目标识别、检测和成像能力，也为政府、垂直行业和用户赋能，其中通感融合架构是其支撑。通智融合以及泛在智能原则包括网络自智和网络的智赋，满足网络的大规模定制、智能供给的需求，其中智能内生架构、数据内生架构、算力内生/云原生架构及数字孪生网络是其支撑。泛在连接/连接未连不仅涉及覆盖能力，还涉及组网生态和商业模式。可持续性除了涉及网络和设备能量效率外，还涉及网络架构和空口制式的演进以及网络的商业模式。安全/隐私/韧性涉及网

基金项目：国家重点研发计划项目（2020YFB1806700）



▲图1 ITU-R定义的IMT-2030典型应用场景

络作为数字和信息基础设施对经济、社会和个人的保障和赋能，同时也涉及对通感融合、通智融合及泛在连接场景的赋能，其中内生安全可信架构是其关键支撑。要灵活满足通信、感知、智能、数据、算力、可信的服务及其灵活编排和能力开放的需求，云原生/算力网络和服务架构是支撑。

中国、欧盟、美国三方是6G网络架构的领导者。从2021年至今，三方都发布了一系列的白皮书和研究报告。以企业的视角对6G网络架构研究进展进行分析，能够为全球企业界、学术界、监管部门等提供参考。本文中，我们将网络架构的设计分为组网生态、先进性能和能效的软硬件联合设计通信架构、云原生/算力网络和服务架构、6G新能力（可信安全内生和智能内生）架构4个维度，并从这4个维度对中国、欧盟、美国三方的观点和研究进展逐一进行介绍和分析。

2 6G网络组网生态研究进展

传统的2G、3G、4G网络组网通常比较固定和封闭。随着5G面向To B千行百业的定制化组网以及天地一体等新组网形态的出现，6G网络的组网生态成为了一种赋能通信产业、垂直行业、社会、国家和个人的新范式。因此，我们将其作为观察中欧美三方6G网络架构研究的第一视角。

2.1 中国6G网络组网生态研究进展

2021年，中国6G研究组织IMT-2030（6G）推进组在6G发展大会上发布了《6G网络架构愿景与关键技术展望》白皮书^[2]，在组网生态方面提出“分布式自治网络”，并将卫星、高空、沉浸多感、确定性通信等子网集成在一起，融入智慧内生、安全内生和数字孪生网络。

2022年，中国IMT-2030（6G）推进组发布《6G分布式网络技术的应用场景及需求研究》报告^[3]，“从业务趋势、网络趋势和网络痛点出发，探讨了6G分布式网络技术的典型应用场景及需求，提出了6G分布式网络的概念、关键特征、要素构成和关键技术需求。”报告指出，分布式网络涉及8项关键技术：云原生、分布式服务框架、分布式智能、分布式数据服务、多方信任框架、分布式近域通信、场景化编排与柔性组装、分布式应用使能。

2.2 欧盟6G网络组网生态研究进展

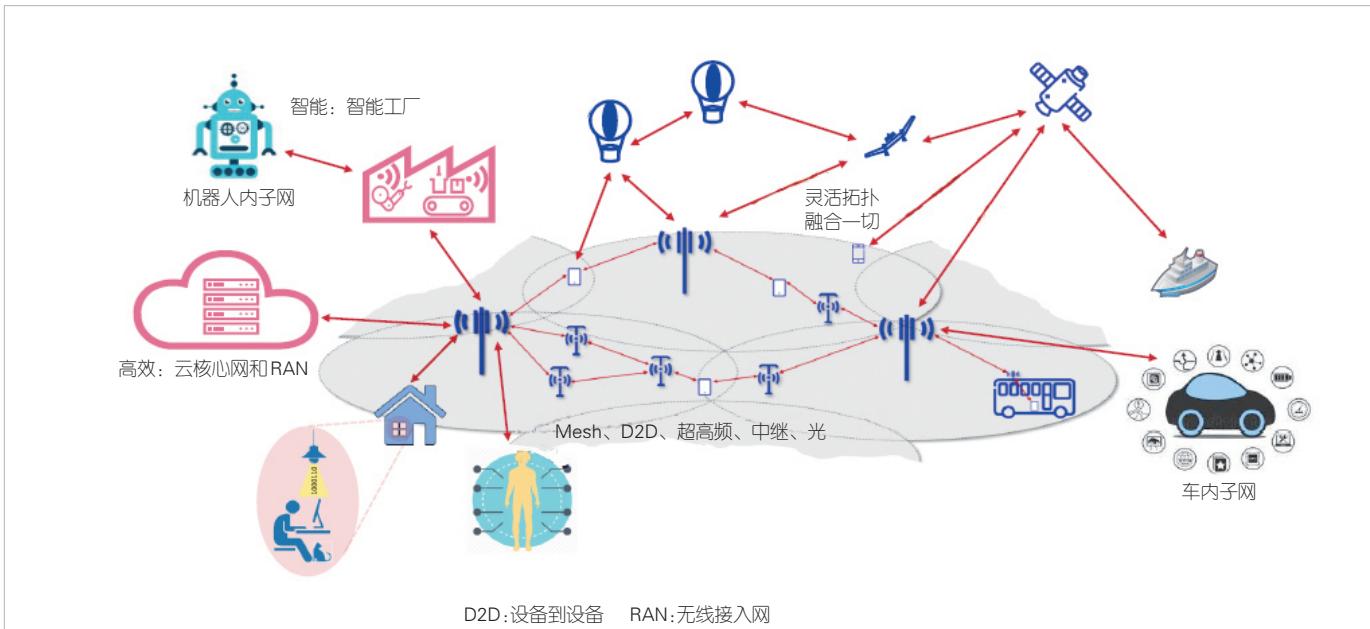
如图2所示，在6G网络研究方面，欧盟启动了6G研究项目HEXA-X，并在2021年底发布《Deliverable D5.1 Initial 6G Architectural Components and Enablers》^[4]，将网络之网作为6G的组网生态。网络之网和中国的分布式自治网络异曲同工，它整合了卫星、汽车、智能工厂等子网，以及Mesh、设备到设备（D2D）、中继等灵活组网方式。在概念上网络之网更侧重不同子网之间的组网，而中国的分布式自治网络更侧重子网的自治。

2.3 美国6G网络组网生态研究进展

和欧盟、中国类似，美国6G研究组织NGA（Next G Alliance）在NextG相关报告^[5-6]中提出先进拓扑和组网方式，包括用户设备（UE）协作、非地面网络（NTN）、极致组网（车内、机器人内、身体内等）、Mesh网络和side-link，以及嵌入式子网等。除此之外，美国在组网韧性方面的重视度高于欧盟和中国，并发布了可信、安全和韧性白皮书^[7]。

3 6G先进性能和能效的软硬件联合设计通信架构研究进展

6G不仅要同时满足高峰值速率（50~200 Gbit/s）、低时延（<1 ms）等极致性能需求，还要满足高灵活性和弹性，以及绿色节能和低成本（高性价比）的需求。由于基于通用硬件（CPU）的软件实现在满足高灵活性和弹性需求的同时，难以满足极致性能和绿色节能/低成本的需求，而基于专用加速器的硬件实现在满足极致性能和绿色节能/



▲图2 欧盟HEXA-X网络之网示意图

低成本的需求时，又难以满足高灵活性和弹性需求，因此要同时满足以上几个方面需求，软硬件联合设计的通信架构是必要的。

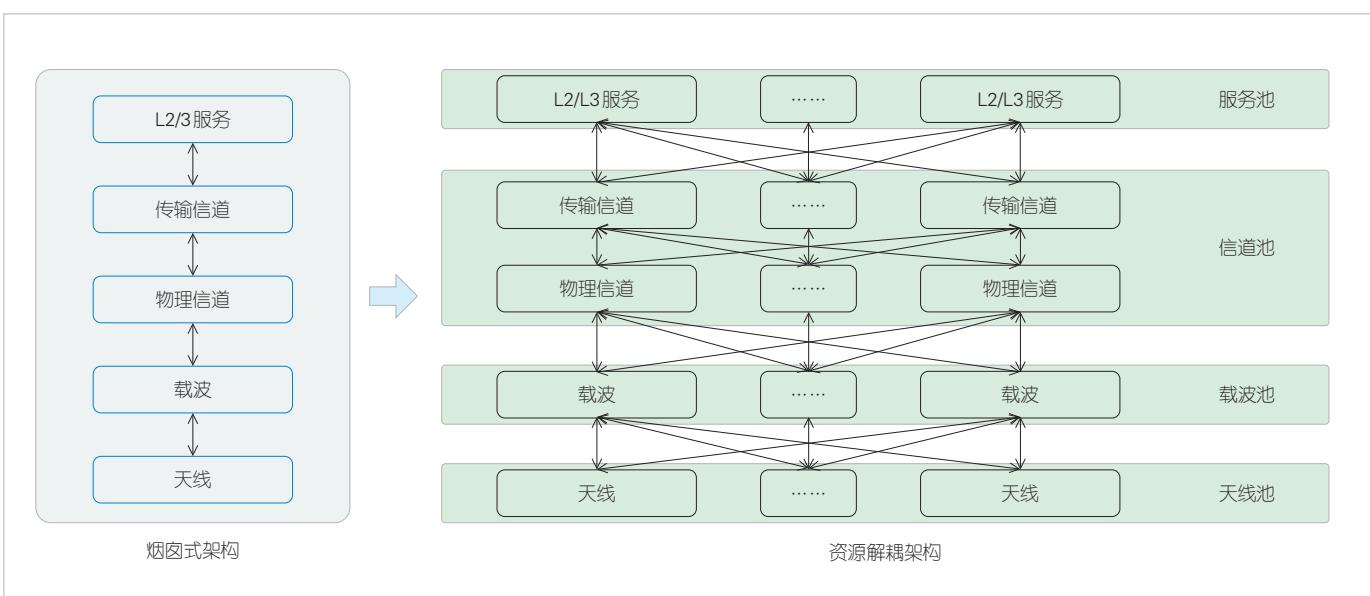
3.1 中国相关研究进展

中国IMT-2030（6G）在先进性能和能效的软硬件联合设计方面开展了几个方面的研究：多频段多发射和接收点（TRP）原生组网、可编排无线接入网（RAN）用户面架构、控制和转发分离RAN架构、控制和转发分离的可编程

网络、确定性网络。

3.1.1 多频段多TRP原生组网

2022年发布的IMT-2030《6G无线网络架构和功能研究报告》^[8]分析了多频段多TRP对空口资源解耦的需求及其影响。如图3左侧所示，当前移动通信蜂窝小区的无线接入网的构建方式是典型的“烟囱式”，表现为：资源与服务紧耦合、上下行传输紧耦合、资源与资源紧耦合。这种紧耦合的方式是一种针对单频段单载波原生的设计，没



▲图3 中国IMT-2030（6G）报告的空口资源解耦全景图

有在一开始就考虑多频段多载波融合组网场景的需求，因而对多频段的支持效率不高，灵活性和按需定制能力不足。这极大限制了移动网络对未来业务场景、需求的适配以及接入网对无线资源的使用效率，还限制了接入网的双向扩展性、易演进性，以及其与AI、大数据等智能化技术的内生式的融合。面向未来的新需求、新技术、新业务和新场景，下一代移动通信蜂窝小区的构建方式，要实现各层之间的“资源解耦”，如图3右侧所示，每层的资源形成“资源池”，每层内的资源之间、层与层的资源之间都满足可灵活编排、可按需定制、可弹性伸缩，以及高性能、高效、低碳等要求。^[8]这实际上是对5G小区架构的重构设计。

3.1.2 可编排RAN用户面架构

IMT-2030（6G）《6G无线网络架构和功能研究报告》分析了Stack-free协议栈和网络智能可编程。其中，Stack-free协议栈涉及功能组件化、数据包矢量化（VPP）、组件并行化、编排灵活化和组件智能化。这意味着现有5G RAN协议栈被重构，以更好地支持转发功能的硬化和软硬件联合实现，在满足先进性能和能效的同时不失灵活性。

3.1.3 控制和转发分离RAN架构

IMT-2030（6G）《6G无线网络架构和功能研究报告》还介绍了支持多TRP/接入点（AP）分布式组网的以用户为中心的接入网架构，采用了控制面和用户面分离的云化控制单元（CCU）和分布式数据单元（DDU）架构，如图4所示。我们认为，CCU/DDU架构是对现有5G集中式单元（CU）/分布式单元（DU）架构的重构，通过控制和转发分离架构更好地满足控制面云化和转发面（用户面）先进性能与能效的差异化要求。

3.1.4 控制和转发分离的可编程网络

在RAN以上的核心网、承载网方面，《6G网络架构愿景与关键技术展望》^[2]介绍了在控制和转发分离的可编程网络方面的思路，从信息通信技术（ICT）融合趋势角度考虑了业界成熟的SDN架构，并阐述了近期快速发展的数据处理单元（DPU）/智能网卡芯片和设备进展。这和5G-A的核心网架构发展趋势一致。

3.1.5 确定性网络架构

《6G网络架构愿景与关键技术展望》^[2]介绍了确定性网

络架构思路，整合了接入网、传输网和核心网的通信保障、感知和测量功能，并通过确定性网络调度与控制中心向确定性服务管理功能提供服务。这个思路和5G-A的发展趋势也基本一致。

3.2 欧盟相关研究进展

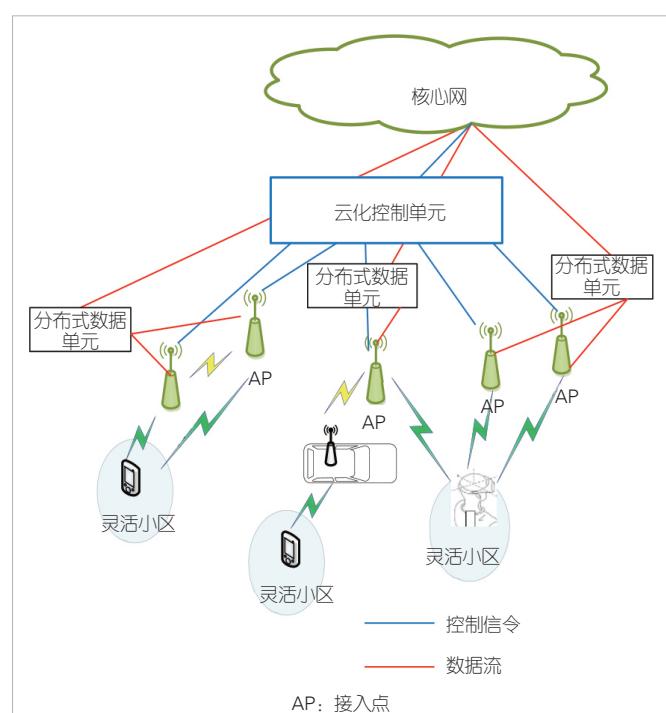
欧盟HEXA-X D5.1^[4]将先进性能和能效的软硬件联合设计主要归入有效网络的研究范畴，但和智能网络也有部分交集。相关主要研究包括RAN功能重构、网络和UE可编程两方面。其中，RAN功能重构与中国的控制和转发分离RAN架构类似，网络可编程与中国的控制和转发分离的可编程网络类似。

3.2.1 多连接

欧盟HEXA-X D5.2^[9]研究了将现有载波聚合/双连接（CA/DC）架构重构为多连接（MC）架构的必要性，汲取CA/DC的长处，支持上下行解耦，以获得极致的可靠性并有效使用频谱。相比于中国IMT-2030（6G）对空口资源解耦的整体设计，欧盟的多连接思路是对5G-A的改进。

3.2.2 用户面协议栈重构

欧盟HEXA-X D5.2^[9]研究了无线链路控制（RLC）协议



▲图4 中国IMT-2030（6G）报告的云化控制单元和分布式数据单元分离架构示意图

栈的ARQ重传以及分组数据汇聚协议（PDCP）按序递交的必要性，并认为这两个功能对TCP传输的性能影响较小，从而使得通过简化协议栈来保持较好的性能成为可能。中国IMT-2030（6G）侧重用户面RAN协议栈可编排的整体性设计，而欧盟的研究则侧重从实证出发，为协议栈的简化和组件化可编排提供支撑。

3.2.3 RAN功能重构

欧盟HEXA-X D5.1^[4]研究了RAN的功能重构，将底层功能（LLF）、用户面的高层功能、控制面功能分离。其中，底层功能需要专用硬件或硬件加速，和射频单元有紧密联系，需要考虑实时性和处理要求，而用户面的高层功能需要适配业务流。

3.2.4 网络和UE可编程

欧盟HEXA-X D5.1^[4]认为，网络和UE可编程将为敏捷交付、灵活性和环境适应性带来价值。智能网卡、软交换机、网络现场可编程门阵列（FPGA）是网络可编程的示例。而可编程UE将为垂直行业专网提供定制化传感器和物联网设备。D5.3^[10]进一步给出了可编程网络的框架和UE可编程的框架。虽然欧盟将网络和UE可编程归入到智能网络的范畴，但我们认为网络和UE可编程也可归入先进性能和能效的软硬件联合设计范畴。

3.2.5 高吞吐低时延

欧盟HEXA-X D5.3^[10]对同时满足高吞吐量(>0.1 Tbit/s)和低时延(<1 ms)要求的架构进行了数值分析，包括对集成的RAN/CN/server架构和拆分的RAN/CN/server架构进行了时延分析。分析结果表明，集成的RAN/CN/server架构的时延优势更加明显。

3.3 美国相关研究进展

美国NGA在先进性能和能效的软硬件联合设计方面的研究，主要涉及多TRP组网^[6]、简化协议栈^[6]、网络解耦^[6]、可编程网络^[11]等方面。

3.3.1 多TRP组网

美国NGA认为，海量分布式多输入多输出（MIMO）通过协同传输和干扰管理，可以有效地提升吞吐量和可靠性^[6]。但该部署方式对架构的影响并未被深入研究。多频段融合对网络协议栈和架构的影响也未在NGA报告中提及。但是NGA考虑了计算资源和能效的联合优化^[5]，认为

分布式计算可通过本地的数据处理和节能，降低数据向远处数据中心传输的能耗，并可利用本地的可持续能源。但分布式本地计算在资源池化增益方面存在不足，使得资源利用率和能效不高，因此需要联合优化计算资源和能效。这个思路和中国的多频段多TRP原生组网的资源池化思想相似。

3.3.2 简化协议栈

美国NGA认为，需要简化协议栈以满足更严格的吞吐量和时延的要求^[6]。物理层和用户面协议栈需要联合设计以缓解综合症，简化缓存管理复杂度，降低开销，动态适应环境变化。报告还认为，空口需要感知级别的服务质量（QoS），应用也需要感知空口条件。相比于中国和欧盟，美国NGA从联合设计的角度提出了一些思路，但尚未给出具体的设计方案。

3.3.3 网络解耦

NGA认为，软硬件解耦能够帮助运营商更灵活地根据需求、技术选择或者供应商，来集成网络^[6]。解耦可以在不同维度上发生：可以在软硬件层面解耦，可以在控制、管理、用户/数据功能之间发生，也可以在RAN的不同层之间发生（例如射频单元/DU/CU）。可以看出，NGA在网络解耦方面并不是主要从性能和能效出发，而是更侧重产业链的重构，其中包括虚拟化/云化、开放化。我们认为，无论怎么虚拟化/云化或者开放化，性能和能效仍是客户关注的关键指标，网络解耦的研究和设计同样需要考虑性能和能效是否有利。

3.3.4 可编程网络

NGA介绍了可编程处理器（可编程网络设备）的研究进展，包括基础设施处理单元（IPU）和DPU^[11]，认为它们可以分担较多的CPU负荷。NGA技术报告^[6]和广域云（WAC）报告^[11]指出，系统级的开放可以使能可编程ICT基础设施。

NGA^[5]认为，传统的基于GPRS隧道协议（GTP）的传输对业务的感知能力很弱，难以适应需要工作链的新的分布式计算时代，因此6G需要对传输协议和相应的控制面功能进行升级。这个方向大致和欧盟以及中国的可编程网络类似。

4 6G云原生、算力网络和服务架构研究进展

面向6G丰富的功能、新能力和业务需求，控制面、业

务面（演进的或IP多媒体子系统）和管理面需要将灵活性、敏捷性、开放性作为设计目标，采用通用软硬件平台。相关研究方向包括云原生、算力网络和服务架构。我们认为，6G的服务架构将从5G时代的从无到有发展到服务架构的全生命周期和生态治理阶段，因此在本文中我们谨慎使用了“服务架构”这个说法，而没有直接沿用“服务化架构（SBA）”这一5G时代的说法。

4.1 中国相关研究进展

《6G分布式网络技术的应用场景及需求研究》^[3]简单介绍了对云原生技术的需求。相比于下文提到的美国NGA在云原生技术和生态体系的研究，中国在云原生以及云和6G融合方面的研究还处于起步阶段。相比而言，中国更期望在算力网络方面发力，构建新的算网融合技术和产业生态。《6G网络架构愿景与关键技术展望》报告^[2]对算力网络初步做了框架分析，包括算力编排管理、算力路由和算力服务能力。

服务架构可以追溯到IT的面向服务架构（SOA）。发展到云原生时代的微服务架构已经成为应用层的一种功能组合和接口架构的框架。5G将服务架构引入到核心网功能和接口定义中，形成了一套有别于SOA和微服务架构风格的SBA。面向6G，《6G分布式网络技术的应用场景及需求研究》^[3]简单介绍了分布式服务基础框架，包括增强的通信负载均衡能力、数据访问代理能力和归一化的高可用组网能力。

2022年IMT-2030 6G无线网络架构报告^[8]初步分析了RAN的服务化需求及其影响。RAN的服务化需要从不同的方面进行具体分析，包括控制面服务化、RAN用户面服务化、多维能力服务化和UE服务化。

4.2 欧盟相关研究进展

欧盟HEXA-X D5.3^[10]将云原生RAN和核心网，以及它们的流水线和重新设计的架构作为有效网络的主要目标，将服务架构延伸到RAN，进一步考虑功能流程的解耦、信令效率的提升、对新AI/机器学习驱动编排的接口支持、计算即服务（CaaS）的框架、基于微服务的SDN控制器等。

4.3 美国相关研究进展

NGA广域云报告^[11]和分布式云和通信系统报告^[5]分别对广域云以及云和通信系统的融合，做了非常全面的分析和研究。从终端设备到接入设备，从站点到城区云、区域

云，都被纳入到广域云的范畴。数据面、通信面和计算面成为自上而下的3层平面。相关报告分析了广域云演进的总体架构、与6G系统的融合、多接入边缘计算（MEC）、设备和网络间/设备间的分布式计算、新的或增强的协议栈、计算和加速的类型、对计算和数据的优化、感知通信的计算编排，以及广域云环境的数据和计算功能的安全隐私。

此外，广域云报告^[11]还初步给出了6G服务架构。UE、RAN、5G功能、演进的5G功能、数据功能、计算功能以及其他6G功能，都被纳入到服务架构中。该架构考虑了编排、负载拆分、服务功能链、服务网格、数据管理、认证和安全，以及相关的Kafka、Istio等成熟技术框架和生态。

5 6G新能力架构研究进展

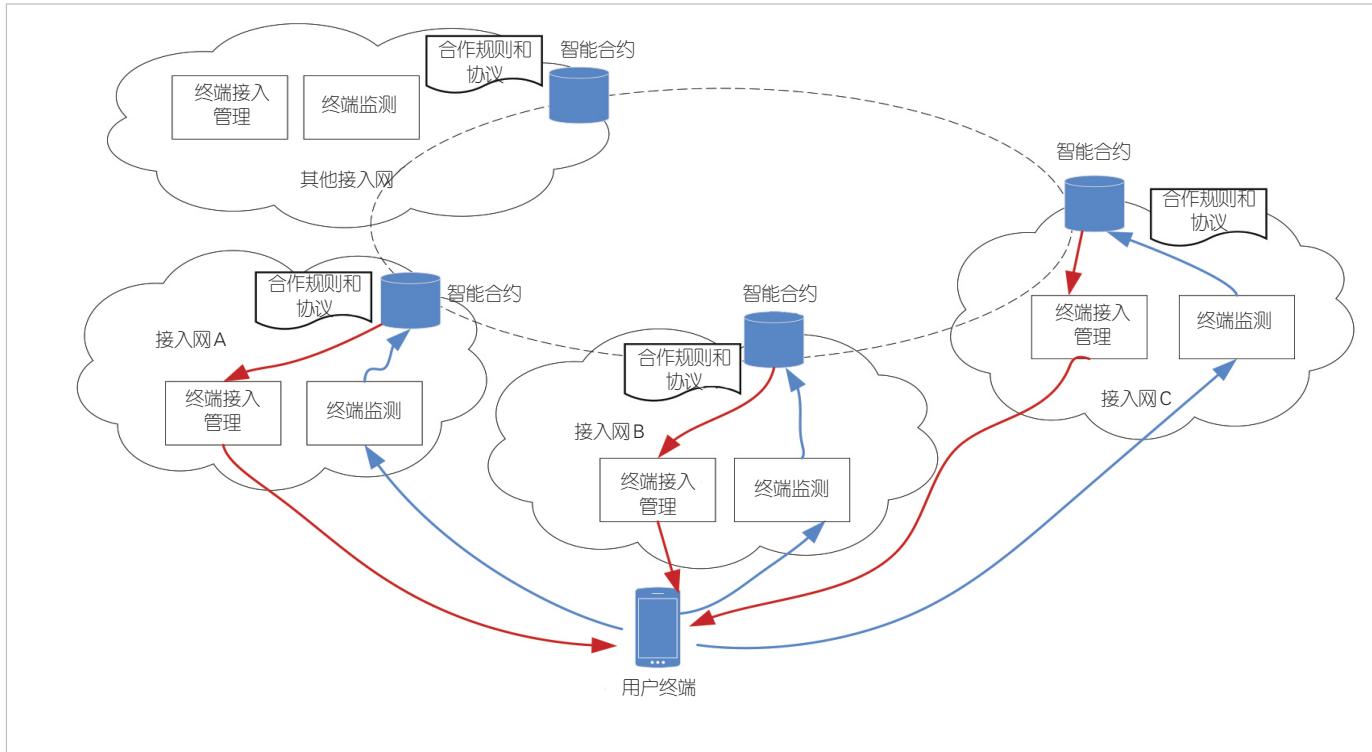
6G新能力主要包括感知、计算、数据、智能和可信五大方面。其中，感知能力和传统的通信、定位能力都涉及空口技术，在架构上也会比较接近通信或定位的架构，因此中国、欧盟、美国三方基本上都没有将感知的架构独立出来进行研究。计算的架构在前文云原生、算力网络和服务架构中已经提到，因此，本章主要涉及数据、智能和可信3个新能力。数据和智能之间有着紧密关系。中欧美三方都将数据和智能联合考虑，并认为智能内生（或原生智能）是6G的特征。因此，我们将数据和智能统一到智能内生架构中。此外，可信和安全的关系也非常紧密，中欧美三方也都把可信安全作为一个合并的研究方向，并且认为可信安全内生是6G的特征。因此，我们也将可信和安全统一到可信安全内生架构中进行介绍。

5.1 中国相关研究进展

5.1.1 可信安全内生架构

IMT-2030（6G）推进组在2021年《6G网络架构愿景与关键技术展望》白皮书^[2]中对分布式网络及其所需的可信即服务（TaaS）做了初步介绍。区块链被引入并内置于网络功能层，以支持多层次链架构。

2022年发布的《6G区块链技术：场景和需求研究》^[12]进一步对区块链的场景、需求和挑战进行了分析。其中，区块链的场景包括存证审计、动态切片管理、动态频谱管理、泛在接入管理、身份/资产管理、分布式安全协作、区块链在协同边缘计算中的应用等方面。在泛在接入管理中，如图5所示，基于区块链的终端接入管理、终端监测和智



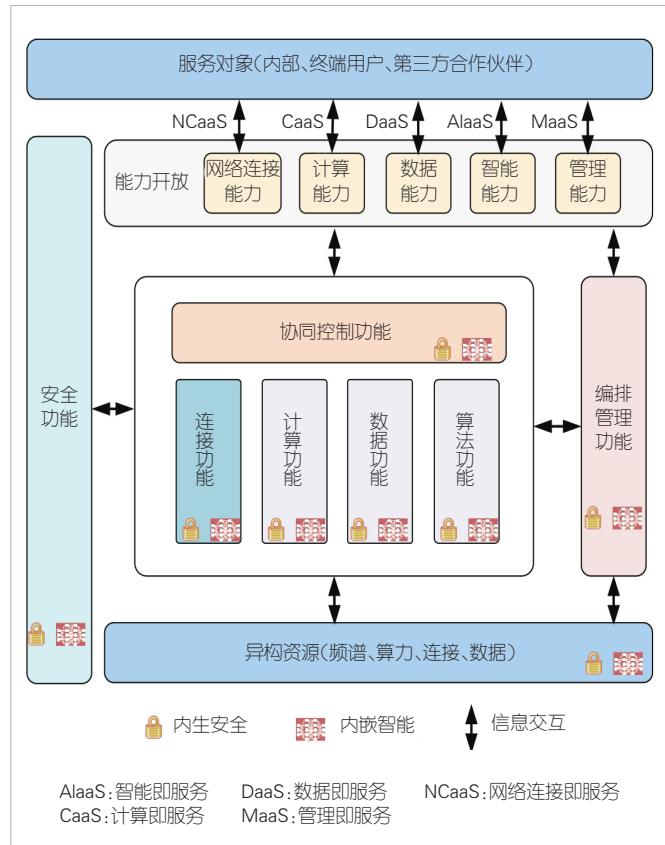
▲图5 中国IMT-2030(6G)报告中基于区块链的泛在接入分布式网络

能合约，将为分布式网络提供泛在接入技术和结算支撑。报告进一步分析了可用于分布式网络中的多链互通、高量区块链、存储扩展等关键技术。

5.1.2 智能内生架构

中国IMT-2030（6G）推进组发布的《6G智能内生网络架构研究》报告对“智能内生”的内涵解释如下：6G智能内生是指，在架构层面将网络连接与人工智能要素（算力、算法、数据）深度融合，以构建网络内完整的智能体系。通过智能的分布式部署和协同，按需对内外提供人工智能即服务（AIaaS），实现智能服务的高效和高质量保障，助力网络实现智能自治，促进智能生态的协同与融合”^[13]。该报告系统分析了6G智能内生的六大特征，具体包括：分层分布式AI，以任务为中心的资源编管控，网络架构、能力和服务的内生自构建，AI与生态协同及融合，AI服务的QoS保障，数智化服务和能力开放。

如图6所示，中国IMT-2030《6G智能内生网络架构研究》报告^[13]进一步对智能内生网络体系框架做了介绍。在资源层面，频谱、算力、连接和数据异构资源进行融合。在功能层面，支撑AI服务的算力、算法、数据、连接四要素首先实现协同控制和融合编排，再通过能力开放服务为网络自身和第三方应用提供AI能力和服务。



▲图6 IMT-2030(6G)报告的智能内生网络体系框架

5.1.3 数据服务架构

2021年发布的《6G网络架构愿景与关键技术展望》白皮书^[2]介绍了可信数据服务的初步框架，其中涉及数据源和数据层（包含数据获取、数据处理、分布式数据存储）。该框架可提供数据即服务（DaaS）给AI能力和应用层。

5.2 欧盟相关研究进展

5.2.1 可信安全内生架构

欧盟在HEXA-X D1.4^[14]展示了安全架构，其中安全可信功能被集成到网络的各个层和网络设备之中，例如RAN、光设备、虚拟化层等。

5.2.2 智能网络和数据服务

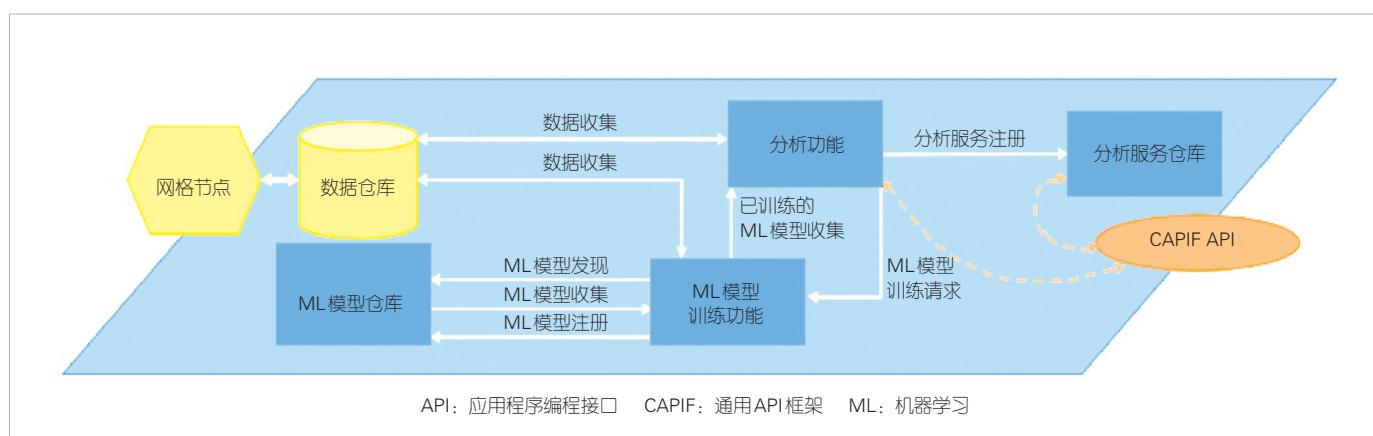
欧盟HEXA-X D5.3^[10]中的智能网络横跨UE、云化RAN/核心网和应用服务器，集成了数据获取、数据开放、

公用的跨域分析（Analytics）框架，以及它们使能的AIaaS和联邦学习即服务（FLaaS），并支持网络自动化和编排，如图7所示。其中，公用的跨域分析框架可以在跨平面和跨域之间交互数据、知识等。AIaaS框架定义了AI训练功能、模型库功能、AI实体功能和AI监控功能，并支持AI模型的发现、存储、更新、训练、推理、性能评估等服务，如图8所示。通过对欧盟和中国对智能网络和数据服务架构的设计，可以看出：欧盟通过分析框架对AI模型和数据做了深度融合设计（即以AI为中心的设计），而中国对AI模型和数据的架构设计则更加独立。

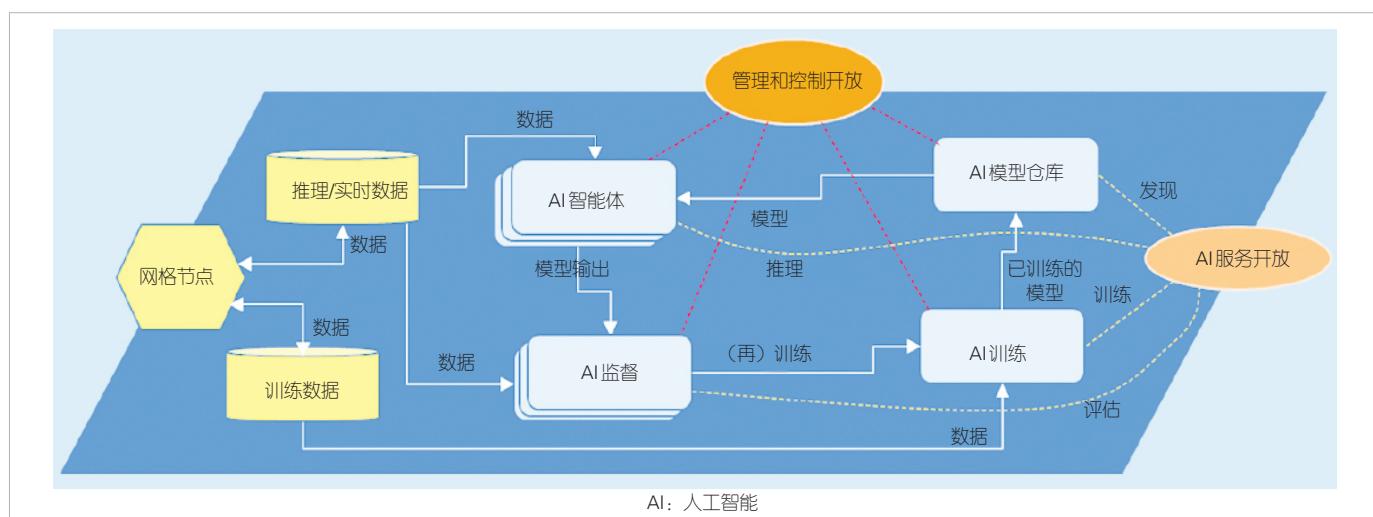
5.3 美国相关研究进展

5.3.1 可信安全内生架构

NGA可信安全韧性白皮书^[7]给出了可信网络的生命周期全景图，包括理解风险和威胁全景，并通过开发、部署和



▲图7 HEXA-X D5.3数据分析框架



▲图8 HEXA-X D5.3人工智能即服务框架

(重) 配置以及运维来应实现, 如图9所示。

5.3.2 智能内生架构

美国 NGA AI 原生无线网络报告^[15]给出了初步的 AI 原生无线网络框架, 并提出了几个研究方向: 1) 集成 AI/ML 到网络中; 2) 支持 AI/ML 的计算架构; 3) 数据集; 4) 安全。报告同时指出该框架面临的挑战。非技术层面最大的困难是数据集的获取。技术层面主要的挑战有: 1) 非确定性的模型需要保底方案, 但这将增加系统复杂性; 2) 分层通信架构和 AI/ML 的集成, 但 AI/ML 更倾向于非分层的架构; 3) 基于电池 UE 的有限资源限制了模型训练和推理的能力。但是 NGA 也表明, 基于美国强大的 AI 领导力, NGA 必将成为无线领域 AI/ML 的领导者和技术推动者。

相比于欧盟和中国, NGA 主要对 AI/ML 对架构的影响做了一些分析, 例如采用边缘计算、分布式计算、集中计算的利弊, 但并未给出任何与架构相关的框架。

6 结语

本文对 2021 年至今中欧美三方在 6G 网络架构方面的研究进展做了综合对比分析。从网络组网生态、先进性能和能效的软硬件联合设计通信架构、云原生/算力网络和服务架构、新能力架构 4 个维度来看, 中欧美三方的架构研

究和设计各有所长, 值得互相借鉴。从中国的研究来看, 在网络组网生态方面可以注意欧盟的网络之网对子网间的组网方面的重视度, 注意美国对韧性的重视度; 在先进性能和能效的软硬件联合设计通信架构方面, 可以学习欧盟和美国从具体实证实例出发研究的方法; 在云原生/算力网络和服务架构方面, 可以思考美国如何将其强大的云原生、算力和服务架构的技术和产业生态为 6G 所用; 在(可信安全内生和智能内生)新能力架构方面, 可以借鉴欧盟的中台式的分析/AI 框架和美国的全生命周期视角的安全可信观点。

6G 网络架构的研究和设计已经从发散逐渐进入到收敛阶段, 期待国内产业界和学术界能够在 6G 网络架构研究中博采众长, 不断精进, 谱写新篇章。

参考文献

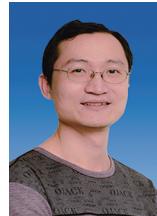
- [1] ITU. Framework and overall objectives of the future development of IMT for 2030 and beyond [EB/OL]. (2023-06-28) [2023-09-15]. <https://www.itu.int/md/R19-SG05-C-0131>
- [2] IMT-2030(6G)推进组. 6G 网络架构愿景与关键技术展望 [R]. 2021
- [3] IMT-2030(6G)推进组. 6G 分布式网络技术的应用场景及需求研究 [R]. 2022
- [4] HEXA-X. Initial 6G architectural components and enablers [EB/OL]. (2021-12-31) [2023-09-15]. https://hexa-x.eu/wp-content/uploads/2022/01/Hexa-X_D5.1_full_version_v1.0.pdf
- [5] NextG Alliance. 6G distributed cloud and communications system [R]. 2022
- [6] NextG Alliance. 6G technologies [R]. 2022
- [7] NextG Alliance. Trust, security, and resilience [R]. 2022



▲图9 NGA可信网络的生命周期全景图

- [8] IMT-2030(6G)推进组. 无线网络架构和功能研究报告 [R]. 2022
- [9] HEXA-X. Analysis of 6G architectural enablers applicability and initial technological solutions [EB/OL]. (2022-12-31)[2023-09-15]. https://hexa-x.eu/wp-content/uploads/2022/10/Hexa-X_D5.2_v1.0.pdf
- [10] HEXA-X. Final 6G architectural enablers and technological solutions [EB/OL]. (2021-12-31)[2023-09-15]. https://hexa-x.eu/wp-content/uploads/2023/05/Hexa-X_D5.3_v1.0.pdf
- [11] NextG Alliance. 6G technologies for wide-area cloud evolution [R]. 2023
- [12] IMT-2030(6G)推进组. 6G 区块链技术: 场景和需求研究 [R]. 2022
- [13] IMT-2030(6G)推进组. 面向6G的智能内生网络架构研究报告 [R]. 2022
- [14] Hexa-X. Hexa-X architecture for B5G/6G networks [EB/OL]. (2021-12-31) [2023-09-15]. https://hexa-x.eu/wp-content/uploads/2023/07/Hexa-X_WP1_D1.4_Summary-slides.pdf
- [15] NextG Alliance. AI-native wireless networks [R]. 2023

作者简介



谢峰，中兴通讯股份有限公司 6G 接入网架构负责人、无线预研首席专家，移动网络和移动多媒体技术国家重点实验室学术带头人；先后参与国家项目5项，获得第8届广东专利奖银奖；发表论文10余篇，申请专利100余项。

综合信息

《中兴通讯技术》2024年专题计划

| 期次 | 专题名称 | 策划人 |
|----|-----------|---|
| 1 | 下一代多址技术 | 北京交通大学教授 艾渤 北京交通大学教授 陈为 |
| 2 | 网络大模型 | 中国电信IP领域首席专家 唐宏 中兴通讯无线首席架构师 熊先奎 |
| 3 | 6G 多天线技术 | 东南大学首席教授、副校长 金石 北京交通大学教授 章嘉懿 东南大学副研究员 韩瑜 |
| 4 | 6G 无线系统技术 | 中国信息通信研究院副院长 王志勤 中国移动研究院院长 黄宇红 东南大学教授 王东明 |
| 5 | 卫星通信技术 | 哈尔滨工业大学(深圳)教授、副校长 张钦宇 |
| 6 | 数据通信新技术 | 中国电信研究院教授级高工 解冲锋 中国联通研究院副院长 唐雄燕 |

面向6G无线环境可预测的感知重构、语义表征及应用



Sensing Reconstruction, Semantic Representation and Application Towards 6G Predictive Wireless Environment

张建华/ZHANG Jianhua, 田艺璇/TIAN Yixuan,
孙语瞳/SUN Yutong, 于力/YU Li

(北京邮电大学 网络与交换技术国家重点实验室, 中国 北京 100876)
(Beijing University of Posts and Telecommunications, State Key Lab of Networking and Switching Technology, Beijing 100876, China)

DOI: 10.12142/ZTETJ.202305007

网络出版地址: <http://kns.cnki.net/kcms/detail/34.1228.TN.20231016.0909.004.html>

网络出版日期: 2023-10-17

收稿日期: 2023-08-10

摘要: 面对6G通信多样复杂的应用场景,精确低复杂的环境信息和信道模型是实现6G智简传输和组网的基础。面向6G无线环境可预测,对无线环境的感知重构、语义表征与应用展开研究,分析了现有的环境感知技术与重构算法,并对无线环境可预测的语义表征方法进行总结。围绕目前存在的信道建模与预测难题,介绍了信道在线预测的6G网络设计与感知重构平台。相关实验验证了该平台的可行性和准确性。认为未来无线环境可预测的关键是提高无线环境感知精度,提升无线环境语义的可解释性,建立一个面向6G信道的通用性系统模型。

关键词: 无线环境; 环境感知; 环境重构; 语义表征; 可预测信道

Abstract: In view of the diverse and complex application scenarios of 6G communications, accurate and low-complexity environment information, and channel models are the basis for achieving the 6G wisdom-evolutionary and primitive-concise network. The sensing reconstruction, semantic representation, and application of wireless environments toward 6G predictive wireless environment are investigated. The existing environment sensing techniques and reconstruction algorithms are analyzed, and the semantic representation methods are summarized for a predictive wireless environment. Focusing on the existing channel modeling and prediction challenges, we introduce the 6G network design method for channel online prediction and the predictive radio environment for digital twin communication platform. The key to future predictive wireless environment is to improve the accuracy of wireless environment sensing, enhance the interpretability of wireless environment semantics, and establish a universal system model for 6G channels.

Keywords: wireless environments; environment sensing; environment reconstruction; semantic representation; predictive channel

引用格式: 张建华, 田艺璇, 孙语瞳, 等. 面向6G无线环境可预测的感知重构、语义表征及应用 [J]. 中兴通讯技术, 2023, 29(5): 38–44. DOI: 10.12142/ZTETJ.202305007

Citation: ZHANG J H, TIAN Y X, SUN Y T, et al. Sensing reconstruction, semantic representation and application towards 6G predictive wireless environment [J]. ZTE technology journal, 2023, 29(5): 38–44. DOI:10.12142/ZTETJ.202305007

第6代移动通信系统（6G）的愿景是提供跨越物理世界和数字世界的即时、高效、智能的通信^[1]。信息论奠基人香农将信道列为通信系统模型的六大模块之一,信道所呈现的特性依赖于电波传播的无线环境及其物理属性。然而,一方面,6G面临Sub-10 GHz、毫米波、太赫兹等多频段,通感一体化^[2]、超高速移动、智能反射面、超大规模多输入多输出（MIMO）等新技术和多样应用的挑战,呈现更

加复杂和高动态的无线环境;另一方面,为了满足未来通信服务高质量、全覆盖的要求,6G网络还面临着复杂多样通信场景带来的巨大挑战^[3-4]。

传统的系统设计方式是先依次得到典型场景的电波传播模型,然后通过离线仿真进行网络设计^[5]。这种方式存在着传输性能远离瞬态最优、环境适应性不足、系统优化成本高、迭代周期长的问题,因此我们迫切需要探索基于物理环境感知重建的自底向上的无线信道预测新范式^[6]。2019年,文献[7]提出基于大数据方法的信道预测框架和思路,并对场景级、米级和波长级的信道特征预测进行分析,进而提出

基金项目: 国家自然科学基金项目 (92167202、61925102、62201086、62101069、62201087)

从模型到预测的范式转变。然而，随着通信频率的升高、传输速率的增加以及散射体材质的多样化，电波在实际环境传播过程中发生反射、折射、衍射时所产生的传播路径和特性变得复杂^[8]。因此，需要对无线环境进行语义表征，并实现信道动态特性的抽象，从而支撑智能化信道预测，如图1所示。

随着人工智能（AI）领域中深度学习（DL）、强化学习（RL）等技术的快速发展，通信系统汇聚了感知和预测的基础能力^[9-11]。一方面，6G网络支持无处不在的连接，通过各种传感器设备，如雷达、位置传感器和摄像头等，能够对物理环境进行实时感知；另一方面，6G的人工智能和计算能力是分析物理环境信息的有力工具，这使得基于感知增强的无线环境预测成为可能。

1 无线环境感知重构

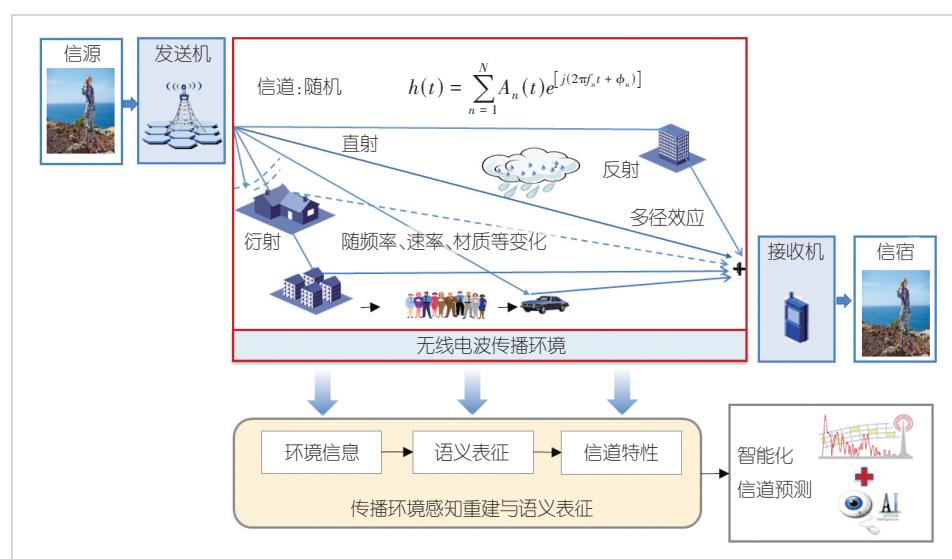
无线环境感知重构即利用传感器和感知技术来获取环境信息，并通过环境重构方法对这些信息进行处理和还原。环境感知是获取环境信息的基础，环境重构的目的则是将感知数据转化为更高层次的环境模型，从而为智能系统的应用和决策提供重要支撑。所以环境感知和三维环境重构是自动驾驶、智能交通系统、机器人导航、数字孪生和虚拟现实等领域的重要组成部分^[12-13]。

1.1 环境感知技术

在复杂高动态的6G无线通信环境中，散射体数量多、

种类繁杂、包含大量环境信息，仅利用单一感知手段难以实现全面、完备的感知。针对不同场景特性，需要结合传感器自身特点加以部署并采集多模态环境数据，从而实现复杂场景下全面的环境感知，支撑6G场景下的无线环境重构。目前的感知设备主要分为视觉和非视觉两种。针对不同设备采集数据的特点，我们在表1中对感知数据类型、设备特点及智能应用进行了总结。

摄像头因具有便捷、成本低等优点成为目前最常用的视觉感知手段之一，能够获取环境中的图像和视频信息。目前主要应用2D计算机视觉技术，如目标检测^[14]、物体跟踪^[15]和语义分割^[16]等方法，对图像中的环境物体和语义信息进行识别和提取。为了感知三维信息，也可以利用多个相机实现视差计算^[17]，对场景的深度信息进行估计，但多设备在三维感知方面的应用目前比较有限。双目相机由两个摄像头集成，不仅可以直接获得左右两个视角的相机图像，实现场景



▲图1 6G通信系统中动态无线环境的语义表征与预测

▼表1 视觉和非视觉感知手段特点及其智能应用

| 手段 | 感知设备 | 感知数据类型 | 设备特点 | 智能应用 |
|-------|--------|-----------|--------------------------------------|----------------|
| 视觉感知 | 摄像头 | 图像、视频 | 部署广泛，丰富的颜色、纹理、形状信息，适用于室内近距离场景、正常光照场景 | 目标检测、物体跟踪、语义分割 |
| | 红外摄像头 | 红外图像、红外视频 | 热像化，有夜视功能，无可见光干扰，适用于低光、夜间等特殊场景 | 目标检测、语义分割、人体识别 |
| | 双目摄像头 | 双视角图像、视频 | 丰富的颜色、纹理、形状、深度信息，适用于室内近距离场景、正常光照场景 | 深度感知、目标跟踪、物体识别 |
| 非视觉感知 | 激光雷达 | 环境点云 | 高精度三维信息，不受光照影响，适用于室外远距离场景，多种光照条件场景 | 三维目标检测、点云语义分割 |
| | 毫米波雷达 | 距离、角度、速度 | 高抗干扰能力、高穿透能力、低功耗，适用于复杂天气条件场景 | 目标检测、距离测量、速度测量 |
| | GPS | 物体位置 | 高精度、全球测量，适用于室外场景 | 定位算法、地图匹配 |
| | 超声波传感器 | 距离测量 | 不受天气影响、成本低，适用于室内、室外近距离场景 | 环境映射、距离测量 |

GPS：全球卫星定位系统

物体的深度感知，还可以提供更准确的目标跟踪和物体识别能力^[18]。红外摄像头弥补了摄像头的光线缺陷，在低光照或无可见光的情况下可以进行目标检测^[19]、语义分割^[20]和人体识别^[21]等处理。

激光雷达是三维感知技术使用的主流传感器，所获取的点云数据可以提供高精度的距离和形状信息，目前主要应用于三维环境重建、三维物体检测^[22]和目标跟踪^[23]等3D计算机视觉技术。此外，激光雷达不受光照条件的影响，可以适用于室外远距离场景和其他各种光线条件场景。但是激光束的穿透能力较差，容易受到浓雾、雨雪天气的影响。毫米波雷达作为经典传感器类型，其感知技术较为成熟，已经被装配于实际的车辆传感中。通过发射毫米波信号来探测物体的距离和速度，毫米波雷达利用目标检测、距离测量、速度测量、障碍物避障等技术从雷达原始信号中提取环境信息^[24-25]。相比激光雷达而言，毫米波雷达信号具有较强的穿透能力，适用于复杂天气条件场景。除此之外，超声波传感器和全球卫星定位系统（GPS）可用于室内、室外场景中目标物体的位置、距离测量，通常应用于物体定位技术。

1.2 三维环境重构技术

在三维环境重构中，运动结构恢复（SfM）、多视点立体视觉（MVS）和点云处理算法是3种主要的技术手段。SfM算法从多视角图像中提取特征点，然后根据这些特征点的匹配关系恢复出相机的运动轨迹和场景的三维结构^[26]。SfM算法生成稀疏点云，适用于静态场景定位，但是在处理大规模场景时容易受到内存和计算资源的限制。MVS算法通过多个视角的图像信息，对场景中的每个像素点进行深度估计^[27]。MVS算法生成稠密点云，可以提供更丰富的场景信息，适用于复杂场景和大规模场景的三维重建。在利用SfM和MVS算法生成环境的点云数据后，使用重建网格^[28]、点云配准^[29]、表面重建^[30]等算法进行点云数据处理，可以生成更精确、完整的三维环境模型。如图2所示，利用双目相机从无线三维环境中获取红绿蓝（RGB）图像和深度图像以构建三维稠密点云图，并采用追踪、局部优化及回环检测进行优化，最终能够生成完整的点云地图。在无线通信系统中，利用环境感知技术和重建方法，感知环境中的障碍物、移动散射体等信息^[6,31]，

对无线环境数据进行实时采集和建模，可以帮助通信系统更好地适应复杂多变的6G通信场景。

2 无线环境语义表征

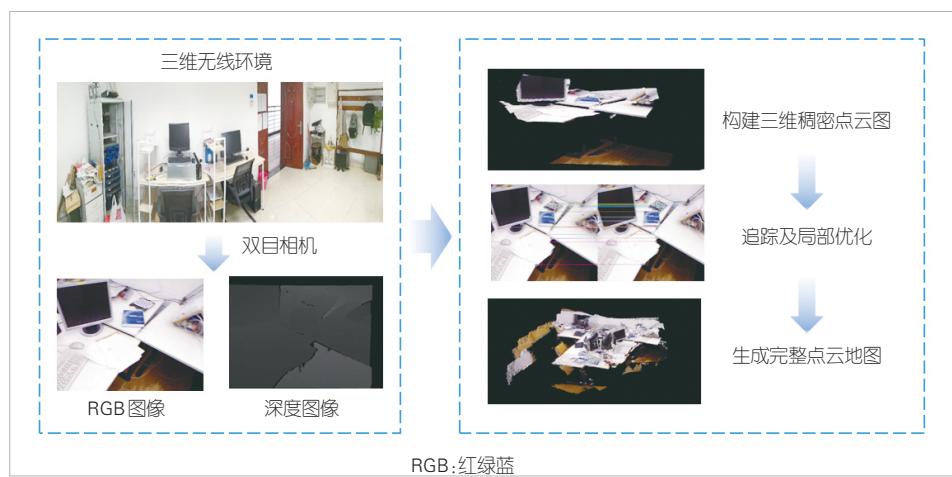
信道传播特性由无线环境决定。无线电波从发射端发出，经与传播路径中的各种散射体发生反射、散射、绕射等相互作用，然后到达接收端。这些接收得到的无线电波会呈现出多径聚簇特征。散射体的形状、位置和材质等都会对电波传播特性产生影响。2016年，文献[32]提出了利用计算机视觉对三维环境进行重构，并定义无线环境散射体与统计簇匹配的簇核（CN）。基于CN的建模方法，我们首先对物理环境的边界以及其中散射体的大小、位置、材质进行重建，然后基于射线追踪（RT）方法实现簇核匹配，赋予簇物理意义^[33]。簇核与信道的匹配关系可以表示为：

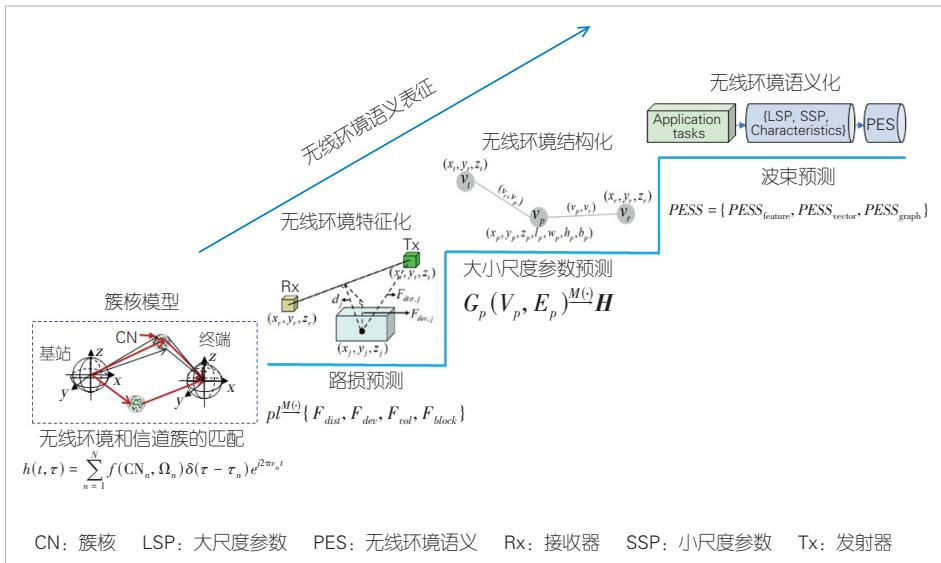
$$h(t, \tau) = \sum_{n=1}^N f(\text{CN}_n, \Omega_n) \delta(\tau - \tau_n) e^{j2\pi v_n t}, \quad (1)$$

其中， $h(t, \tau)$ 表示信道系数， $f(\text{CN}_n, \Omega_n)$ 表示CN与散射体之间的映射关系， $\delta(\tau)$ 表示时延 τ 的狄拉克函数， v_n 表示第 n 个CN的多普勒频移。

为了进一步研究无线环境与信道的匹配映射关系，我们对无线环境进行特征化^[34]、结构化^[35]、语义化^[36]，实现对无线环境的解构和表征，并进一步利用不同无线传播的特征实现信道预测，如图3所示。

我们首先对无线环境进行特征化解构，提出了一种基于环境特征的路损预测模型^[34]。该预测模型用散射体距离、偏移、体积以及直射径（LOS）遮挡情况4种低维特征来表征环境，并基于所提环境特征利用随机森林算法进行路损预测。路损和特征的映射关系可以表示为：





▲图3 无线环境语义表征

$$pl^{M(\cdot)}\{F_{dist}, F_{dev}, F_{vol}, F_{block}\}, \quad (2)$$

其中， pl 为路径损耗， $M(\cdot)$ 为映射函数， F_{dist} 、 F_{dev} 、 F_{vol} 、 F_{block} 分别为距离、偏移、体积、直射径遮挡情况4个环境特征。

仿真结果表明，相比于基于人工神经网络(ANN)和卷积神经网络(CNN)方法的模型，该模型可以提供与基于RT方法接近的结果，并且不同特征组合可以适用于环境信息获取和预测精度的不同要求。

不同于大尺度参数与全局环境特征相关这一特征，小尺度参数对环境的布局更加敏感。为了实现无线环境的结构化表征，我们利用图结构数据对无线环境进行抽象，提出了一种基于图神经网络(GNN)的信道预测方法^[35]。该方法首先以散射体为中心构建通信环境图(CEG)，通过实现有效散射体(参与传播的散射体)以及一次传播散射体进行识别；然后根据识别结果构建相应的无线环境图(PEG)，并基于GNN分别实现大尺度参数(LSP)和小尺度参数(SSP)预测。用 $G_p(V_p, E_p)$ 表示具有节点 V_p 、边 E_p 和节点特征向量 $X_v(v \in V_p)$ 的传播环境图。则环境图结构与信道的映射关系可以表示为：

$$G_p(V_p, E_p)^{M(\cdot)} H, \quad (3)$$

其中， $M(\cdot)$ 为映射函数， H 为信道矩阵。

为了进一步解决无线环境高维度信息如何解构，以及其与信道参数的复杂映射机理难题，我们对无线环境进行了语义化表征，提出了一种以任务为导向的通信任务-信道特性-无线环境的3层解构框架，明确了无线环境语义(PES)

的概念，指出PES是影响无线传播特性(衰落)的主要环境变量，可以采用无线环境语义符号(PESS)作为基本单元进行表示^[36]。PESS包括特征 $PESS_{feature}$ 、线性向量 $PESS_{vector}$ 以及非线性图 $PESS_{graph}$ ：

$$PESS = \{PESS_{feature}, PESS_{vector}, PESS_{graph}\} \quad (4)$$

在这项工作中，以信道特性为桥梁，利用任务-信道以及信道-环境的先验知识，给出任务导向的无线环境语义构建方法。具体步骤包括：首先，针对通信任务需求相关的信道特性，包括LSP、SSP、LOS遮挡特性，将无线环境从全局以及

局部维度进行解构；然后，分别针对全局以及局部环境信息，进行环境结构化和特征化表征；最后，以波束预测任务为例验证所提方法性能，针对信道质量评估以及波束预测分别给出无线环境语义的构建方法，并基于机器学习算法部署预测模型。研究结果表明，相比于目前的基于环境图像的信道预测方法，所提方法精度至少提高6%，并且预测时间降低一个数量级。

3 无线环境可预测

目前环境感知增强在无线通信系统中的应用受到了广泛的关注，如信道大小尺度参数预测、波束赋形、遮挡预测、天线选择、资源管理、干扰控制等。通过感知环境中的物理参数、障碍物、移动散射体等信息，为信道预测算法提供可靠的输入数据，帮助系统预测信道的动态变化和多径传播特性。基于由簇核建模到基于AI的信道预测的研究方法，针对无线通信网络面临的复杂多变的无线环境和信道问题，文献[6]从电波传播角度，提出了一种6G可预测网络——EWaveNet，即利用增强感知实现物理无线环境的重构。EWaveNet设计了与传统网络不同的包括感知、预测和决策三大功能的预测平面，并分析设计了网络架构在跨层优化、隔离性和灵活性方面的优势。其中的环境感知增强模块可以利用普通相机、深度相机、激光雷达等多种设备实现多模态数据采集，并利用特征工程、特征学习相关技术提取环境的低级基本、中级隐性和高级语义的多维度特征，建立完备的三维环境数据集。

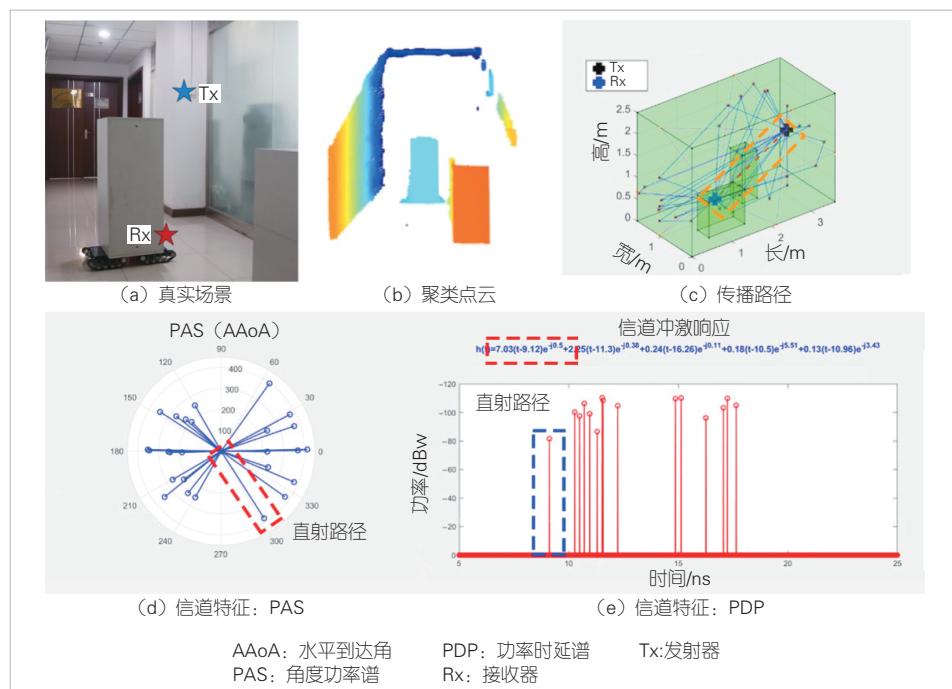
进一步地，预测模块从无线电波传播机理出发，通过感

知物理环境来建立相应的三维重构环境。基于簇核匹配和电场计算方法，利用由多维环境特征和信道数据构成的环境数据集，实现信道参数和特性的快速预测。决策模块利用预测结果支撑通信传输算法设计及组网优化，例如实现信道链路的提前切换，预判用户移动轨迹上的遮挡并提前选择最优波束，支持高速移动、高低频融合等场景下的空-时-频信道推演，简化信道估计，减少反馈开销等。

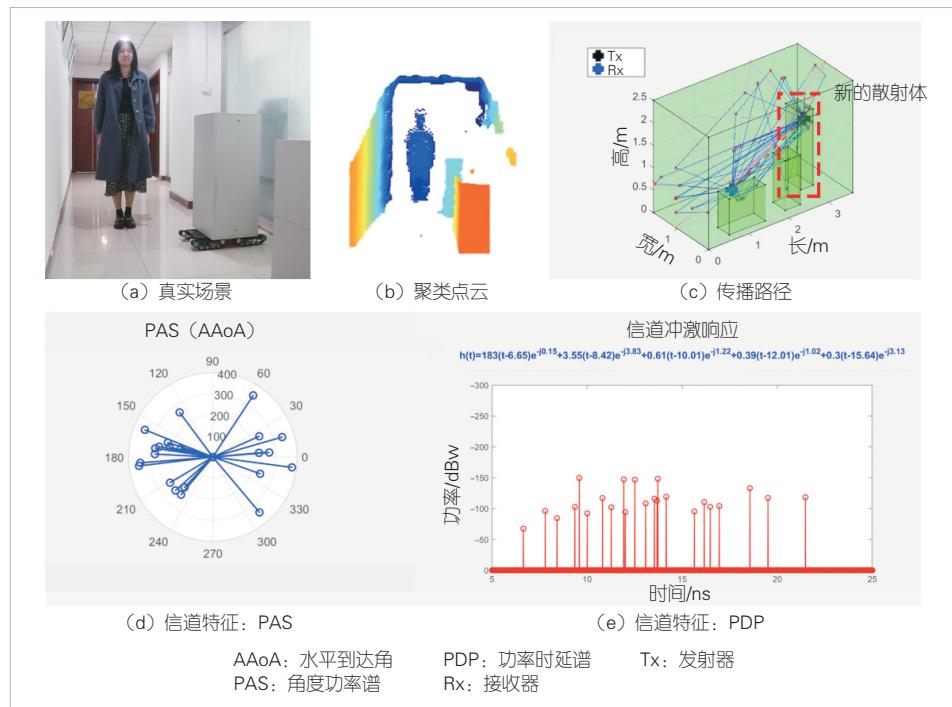
EWaveNet 创新性地实现了一种自底向上的物理环境感知、信道信息预测、智能传输设计和组网性能优化的孪生网络，在此基础上，文献[37]搭建了感知增强的无线环境预测与重建平台——PREDICT-Plat。该平台由 3 个模块组成：环境感知模块、电波传播预测模块、数据处理和可视化模块。首先，环境感知模块通过引入计算机视觉技术，利用传感设备实时采集环境信息并生成点云数据，配合特征提取计算散射体的位置、尺寸等信息，实现环境散射体的实时采集与自动重建；然后，电波传播预测模块采用优化的超快速、高精度的电场计算方法和环境信道关联数据库，在重建的环境中进行电磁计算，以获取电波传播路径信息，包括传播方向、反射和衍射，并且计算路径参数，例如幅度、时延和角度；最后，数据处理和可视化模块实现时变场景下信道衰落实时预测和展示，直观地呈现当前环境下的信道特性，如信道冲激响应（CIR）、功率时延谱（PDP）和角度功率谱（PAS）。平台通过时变场景快速感知和动态电磁仿真技术，可高精度模拟推演无线环境变化（散射体运动）规律。

该平台在 28 GHz 室内场景下通过了视距/非视距、有人/无人情况的实测验证。视距传播场景和遮挡场景预测结果分别如图 4 和图 5

所示。当散射体实时移动和人员走动时，平台能通过感知设备快速捕捉环境变化，生成聚类点云并完成环境重构，随后通过预测模块实时更新信道衰落预测结果，从而验证了平台的可行性和准确性。该平台解决了复杂应用场景下环境感知数据的快速实时处理和多维多尺度无线环境预测难题，有望自底向上地提高 6G 网络对环境变化的适应能力，是面向 6G



▲图 4 视距场景的信道衰落预测结果^[33]



▲图 5 存在行人的信道衰落预测结果^[33]

无线环境“数字孪生”的初步形态。但是目前该平台仍受到采集设备精度和电磁计算效率等因素的限制。未来，在更先进的传感器和强大计算能力的支持下，该平台可以完成6G网络多种场景下无线环境的重构与预测。

4 未来挑战与展望

1) 提高环境感知的精度

目前感知技术受到传感器设备精度、采集数据质量、算法复杂度和实时性等多方面的限制，这导致最终重构出的环境也存在一定误差。同时感知算法的高复杂性，对计算设备的内存和算力提出了较大的挑战。提升环境感知技术的精度需要从传感器技术、数据质量、算法改进、实时性优化等多个方面入手。提高感知设备和算法在各种应用场景下的准确性和稳定性，获取丰富的信息，有助于实现较高的无线环境感知精度。

2) 提升无线环境语义可解释性、泛化性

现有信道预测研究大多使用深度学习技术进行信道下游任务的预测，然而深度学习模型由大量的神经元和权重组成，导致其决策过程难以直观理解，通常在语义可解释性方面存在挑战。同时机器学习模型在测试数据或实际应用中表现较差，相关的应用场景及数据较为固定。提高语义可解释性和泛化性是不断探索和优化的方向，这使得机器学习模型更可靠、更具实用性，从而提高无线环境语义表征的场景适应性。

3) 普适化无线环境的可预测框架

不同的通信场景和下游任务需要不同类型的感知设备来进行信息采集。文献[38]对基于完整环境图像及基于语义目标图像的波束预测方法进行了对比。结果表明，基于语义目标的方法能有效去除环境冗余信息，增强预测模型的鲁棒性。现有的基于环境信息进行信道预测的工作大多为任务导向，面向不同的下游信道预测任务进行信道预测，缺乏一个通用性的统一架构。因此，如何设计一个大而统的无线环境可预测框架也是一个亟待解决的难题。

5 结束语

到目前为止，虽然自动驾驶领域的环境感知重构技术已经基本成熟，但是无线环境可预测仍处于研究的初级阶段。本文面向6G网络中信道复杂动态变化的难题，总结了现有的感知重构、语义表征与信道可预测的研究进展。然而，现有研究存在感知技术精度较低、无线环境语义可解释性和泛化性较差、环境可预测受到任务限制等问题。如何提高无线环境感知精度，提升无线环境语义的可解释性，建立一个面

向6G信道的通用性系统模型，是今后研究的关键所在。

参考文献

- [1] ZHANG P, XU W J, GAO H, et al. Toward wisdom–evolutionary and primitive–concise 6G: a new paradigm of semantic communication networks [J]. Engineering, 2022, 8: 60–73. DOI: 10.1016/j.eng.2021.11.003
- [2] ZHANG J H, WANG J L, ZHANG Y X, et al. Integrated sensing and communication channel: measurements, characteristics, and modeling [J]. IEEE communications magazine, 2023, (99): 1–7. DOI: 10.1109/MCOM.020.2300165
- [3] LIU G Y, LI N, DENG J, et al. The SOLIDS 6G mobile network architecture: driving forces, features, and functional topology [J]. Engineering, 2022, 8: 42–59. DOI: 10.1016/j.eng.2021.07.013
- [4] ZHANG J H, LIN J X, TANG P, et al. Channel measurement, modeling, and simulation for 6G: a survey and tutorial [EB/OL]. (2023–05–26) [2023–08–08]. <https://arxiv.org/abs/2305.16616.pdf>
- [5] 张建华, 唐盼, 姜涛, 等. 5G信道建模研究的进展与展望 [J]. 中国科学基金, 2020, 34(2): 163–178. DOI: 10.13443/j.cjors.2017091505
- [6] NIE G F, ZHANG J H, ZHANG Y X, et al. A predictive 6G network with environment sensing enhancement: from radio wave propagation perspective [J]. China communications, 2022, 19(6): 105–122. DOI: 10.2391/JCC.2022.06.009
- [7] LI W, ZHANG J H, MA X C, et al. The way to apply machine learning to IoT driven wireless network from channel perspective [J]. China communications, 2019, 16(1): 148–164. DOI: 10.12676/j.cc.2019.01.014
- [8] ZHANG J H, TANG P, YU L, et al. Channel measurements and models for 6G: current status and future outlook [J]. Frontiers of information technology & electronic engineering, 2020, 21(1): 39–61. DOI: 10.1631/FITEE.1900450
- [9] 刘留, 张建华, 樊圆圆, 等. 机器学习在信道建模中的应用综述 [J]. 通信学报, 2021, 42(2): 134–153. DOI: 10.11959/j.issn.1000–436x.2021001
- [10] ZHANG Z, ZHANG J H, ZHANG Y X, et al. AI-based time-, frequency-, and space-domain channel extrapolation for 6G: opportunities and challenges [J]. IEEE vehicular technology magazine, 2023, 18(1): 29–39. DOI: 10.1109/MVT.2023.3234169
- [11] ZHANG Z, ZHANG J H, ZHANG Y X, et al. Deep reinforcement learning based dynamic beam selection in dual-band communication systems [J]. IEEE transactions on wireless communications, 2023, (99): 1. DOI: 10.1109/TWC.2023.3300830
- [12] 刘光毅, 邓娟, 郑青碧. 基于数字孪生网络的6G无线网络自治 [J]. 中兴通讯技术, 2023, 29(3): 2–7. DOI: 10.12142/ZTETJ.202303002
- [13] 段向阳, 康红辉, 吕星哉, 等. 面向6G的无线接入网络数字孪生技术 [J]. 中兴通讯技术, 2023, 29(3): 32–37. DOI: 10.12142/ZTETJ.202303007
- [14] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection [C]//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2016: 779–788. DOI: 10.1109/CVPR.2016.91
- [15] WOJKE N, BEWLEY A, PAULUS D. Simple online and realtime tracking with a deep association metric [C]//Proceedings of 2017 IEEE International Conference on Image Processing (ICIP). IEEE, 2018: 3645–3649. DOI: 10.1109/ICIP.2017.8296962
- [16] CHEN L C, PAPANDREOU G, KOKKINOS I, et al. DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs [J]. IEEE transactions on pattern analysis and machine intelligence, 2018, 40(4): 834–848. DOI: 10.1109/TPAMI.2017.2699184
- [17] CHEN Y L, LIU S, SHEN X Y, et al. DSGN: deep stereo geometry network for 3D object detection [C]//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2020: 12533–12542. DOI: 10.1109/CVPR42600.2020.01255
- [18] LAGA H, JOSPIN L V, BOUSSAID F, et al. A survey on deep learning techniques for stereo-based depth estimation [J]. IEEE transactions on pattern analysis and machine intelligence, 2022, 44(4): 1738–1764. DOI: 10.1109/TPAMI.2020.3032602
- [19] MUKHERJEE S, COUDERT O, BEARD C. UNIMODAL: UAV–aided infrared imaging based object detection and localization for search and disaster recovery [C]//Proceedings of 2022 IEEE International Symposium on Technologies for Homeland Security (HST). IEEE, 2023: 1–6. DOI: 10.1109/HST56032.2022.10025436
- [20] XU Z G, WANG J, WANG L Y. Infrared image semantic segmentation based on improved DeepLab and residual network [C]//Proceedings of

- 2018 10th International Conference on Modelling, Identification and Control (ICMIC). IEEE, 2018: 1–9. DOI: 10.1109/ICMIC.2018.8530003
- [21] KOPACZKA M, SCHOCK J, NESTLER J, et al. A combined modular system for face detection, head pose estimation, face tracking and emotion recognition in thermal infrared images [C]//Proceedings of 2018 IEEE International Conference on Imaging Systems and Techniques (IST). IEEE, 2018: 1–6. DOI: 10.1109/IST.2018.8577124
- [22] CHARLES R Q, HAO S, MO K C, et al. PointNet: deep learning on point sets for 3D classification and segmentation [C]//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2017: 77–85. DOI: 10.1109/CVPR.2017.16
- [23] XIANG Z H. A simultaneous object detection and tracking framework based on point cloud [C]//Proceedings of 2022 14th International Conference on Computer Research and Development (ICCRD). IEEE, 2022: 225–229. DOI: 10.1109/ICCRD54409.2022.9730356
- [24] GAO X Y, XING G B, ROY S, et al. RAMP-CNN: a novel neural network for enhanced automotive radar object recognition [J]. IEEE sensors journal, 2021, 21(4): 5119–5132. DOI: 10.1109/JSEN.2020.3036047
- [25] MEYER M, KUSCHK G, TOMFORDE S. Graph convolutional networks for 3D object detection on radar data [C]//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW). IEEE, 2021: 3053–3062. DOI: 10.1109/ICCVW54120.2021.00340
- [26] DING L, ELLIETHY A, SHARMA G. 3D georegistration of wide area motion imagery by combining SFM and chamfer alignment of vehicle detections to vector roadmaps [C]//Proceedings of 2017 IEEE International Conference on Image Processing (ICIP). IEEE, 2018: 1487–1491. DOI: 10.1109/ICIP.2017.8296529
- [27] BERRA E F, PEPPA M V. Advances and challenges of uav sfm mvs photogrammetry and remote sensing: short review [C]//Proceedings of 2020 IEEE Latin American GRSS & ISPRS Remote Sensing Conference (LAGIRS). IEEE, 2020: 533–538. DOI: 10.1109/LAGIRS48042.2020.9285975
- [28] LUO Q, LI Y, QI Y. Distributed refinement of large-scale 3D mesh for accurate multi-view reconstruction [C]//Proceedings of 2018 International Conference on Virtual Reality and Visualization (ICVRV). IEEE, 2019: 58–61. DOI: 10.1109/ICVRV.2018.00018
- [29] WANG H Q, HUANG L, YU K, et al. Deep-learning-based multiview RGBD sensor system for 3-D face point cloud registration [J]. IEEE sensors letters, 2023, 7(5): 1–4. DOI: 10.1109/LSENS.2023.3267948
- [30] WANG Z Y, ISLER V, LEE D D. Surface Hof: surface reconstruction from a single image using higher order function networks [C]//Proceedings of 2020 IEEE International Conference on Image Processing (ICIP). IEEE, 2020: 2666–2670. DOI: 10.1109/ICIP40778.2020.9190842
- [31] ALKHATEEB A, CHARAN G, OSMAN T, et al. DeepSense 6G: a large-scale real-world multi-modal sensing and communication dataset [J]. IEEE communications magazine, 2023, (99): 1–7. DOI: 10.1109/MCOM.006.2200730
- [32] ZHANG J H. The interdisciplinary research of big data and wireless channel: a cluster-nuclei based channel model [J]. China communications, 2016, 13(Supplement2): 14–26. DOI: 10.1109/CC.2016.7833457
- [33] YU L, ZHANG Y X, ZHANG J H, et al. Implementation framework and validation of cluster-nuclei based channel model using environmental mapping for 6G communication systems [J]. China communications, 2022, 19(4): 1–13. DOI: 10.23919/JCC.2022.04.001
- [34] SUN Y T, ZHANG J H, ZHANG Y X, et al. Environment features-based model for path loss prediction [J]. IEEE wireless communications letters, 2022, 11(9): 2010–2014. DOI: 10.1109/LWC.2022.3192516
- [35] SUN Y T, ZHANG J H, ZHANG Y X, et al. Environment information-based channel prediction method assisted by graph neural network [J]. China communications, 2022, 19(11): 1–15. DOI: 10.23919/JCC.2022.11.001
- [36] SUN Y T, ZHANG J H, YU L, et al. How to define the propagation environment semantics and its application in scatterer-based beam prediction [J]. IEEE wireless communications letters, 2023, 12(4): 649–653. DOI: 10.1109/LWC.2023.3237827
- [37] MIAO Y H, ZHANG Y X, ZHANG J H, et al. Demo abstract: predictive radio environment for digital twin communication platform via enhanced sensing [C]//Proceedings of IEEE INFOCOM 2023–IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS). IEEE, 2023: 1–2. DOI: 10.1109/INFOCOMWKSHPS57453.2023.10225977
- [38] SUN Y T, ZHANG J H, WANG J L, et al. PC-SC: a predictive channel-based semantic communication system for sensing scenarios [J]. Electronics, 2023, 12(14): 3129. DOI: 10.3390/electronics12143129

作者简介



张建华,北京邮电大学教授、理学院副院长、博士生导师,北邮-中国移动联合创新中心主任;研究方向为6G移动通信技术、人工智能、数据挖掘,主要包括智能信道建模、大规模MIMO和太赫兹信道建模、信道仿真仪和OTA测试;发表论文300余篇,授权专利50余项。



田艺璇,北京邮电大学在读硕士研究生;主要研究方向为信道预测和环境感知。



孙语瞳,北京邮电大学在读博士研究生;主要研究方向为信道预测、环境感知、计算机视觉。



于力,北京邮电大学博士后;主要研究方向为智能信道建模、信道预测和机器学习。



面向下一代移动通信的 前传网络关键技术

Key Technologies for Fronthaul Network in Next-Generation Mobile Communication

赵俊皓/ZHAO Junhao, 张俊文/ZHANG Junwen,
迟楠/CHI Nan
(复旦大学 电磁波信息科学教育部重点实验室, 中国 上海 200433)
(Key Laboratory of EMW Information (MoE), Fudan University, Shanghai 200433, China)

DOI: 10.12142/ZTETJ.202305008
网络出版地址: <http://kns.cnki.net/kcms/detail/34.1228.TN.20231016.1012.010.html>
网络出版日期: 2023-10-17
收稿日期: 2023-08-08

摘要: 下一代宽带移动通信的容量相对5G通信提升百倍至千倍,这对支撑其发展的前传光通信网络的物理基础提出了巨大的挑战。系统回顾和梳理了移动前传网络的需求、技术和实现架构等。以光纤无线融合接入为基础,重点针对数字前传技术、模拟前传技术和数模结合前传技术3类移动前传架构,围绕其网络架构、关键技术和未来发展方向进行了深入的分析和探讨,并对未来前传网络关键技术进行展望。本研究可为未来移动前传网络的研究提供参考。

关键词: 前传网络; 接入网; 光载无线

Abstract: The next-generation broadband mobile communication is expected to increase capacity by hundreds to thousands of times compared with 5G communication, posing significant challenges to the physical foundation of the fronthaul optical communication network that supports its development. A systematic review of mobile fronthaul networks is provided, including their requirements, technologies, and implementation architectures. Furthermore, three types of mobile fronthaul architectures based on radio-over-fiber access are highlighted: digital fronthaul technology, analog fronthaul technology, and hybrid digital-analog fronthaul technology. Then the network architectures, key technologies, and future development directions are discussed, offering valuable insights and references for the research on the mobile fronthaul networks of future communication.

Keywords: fronthaul network; access network; radio-over-fiber

引用格式: 赵俊皓, 张俊文, 迟楠. 面向下一代移动通信的前传网络关键技术 [J]. 中兴通讯技术, 2023, 29(5): 45–55. DOI: 10.12142/ZTETJ.202305008

Citation: ZHAO J H, ZHANG J W, CHI N. Key technologies for fronthaul network in next-generation mobile communication [J]. ZTE technology journal, 2023, 29(5): 45–55. DOI: DOI: 10.12142/ZTETJ.202305008

进入21世纪以来,通信需求的爆炸性增长给宽带通信接入带来巨大考验,也为新型高速通信技术的发展提供机遇。如图1(a)所示,与5G相比,下一代宽带移动通信在数据容量、网络覆盖范围、能源消耗、时延、可靠性以及网络连接密度方面提出新的需求^[1],通信容量提升百倍至千倍,这对支撑其发展的前传光通信网络的物理基础提出重大挑战。随着数据流量的指数级增长,现有前传光网络带宽已不能完全满足未来业务流量需求,亟需物理层通信技术的变

革。当前,以通用公共无线接口(CPRI)为主要接口标准的集中式无线接入网^[2]和云无线接入网是主流的前传网络。但由于存在低频谱效率和大数据速率需求,CPRI的可扩展性受到限制。如今,移动前传网络的发展面临两大挑战:一方面,未来移动前传网络亟需寻求新的光前传网络实现架构,实现大带宽、高速率、低时延的移动前传网络;另一方面,亟需开发新型前传数据调制、传输和解调实现方式,压缩前传通信的带宽,以进一步提高前传通信的传输效率。

对此,业界展开了大量深入的研究。在网络架构实现方面,相干技术的下沉为高速短距离光接入及光互连提供了机遇,包括基于相干载波聚合技术的点到多点相干技术^[3],以及基于相干的无源光网络^[4],均有望提供巨大的前传通信网

基金项目: 国家自然科学基金资助项目(62171137、62235005、61925104);上海市自然科学基金资助项目(21ZR1408700)

络带宽，这也为超高速的移动前传光网络提供了物理基础。另一方面，随着光纤无线电融合（RoF）技术的发展，新型前传数据调制、传输和解调成为可能。针对CPRI低效率的数据传输，压缩前传网络的数据量成为研究热点^[5]。有望替代CPRI的前传技术分支有：数字前传技术、模拟前传技术和数模结合前传技术等。其中，数字前传技术的研究热点聚焦于压缩数字信号数据，以获取更高频谱效率^[6]；模拟前传技术的研究热点在于如何更好地补偿模拟信号的非线性失真^[7]；数模结合前传技术则结合了数字信号的高保真度和模拟信号的高频谱效率的优点，实现了两者间较好的结合^[8]。而近年来，随着毫米波、太赫兹等无线通信技术的发展，业界也提出了光子毫米波、光子太赫兹承载的前传网络^[9-10]，以满足未来前传网络的大规模密集布站需求。

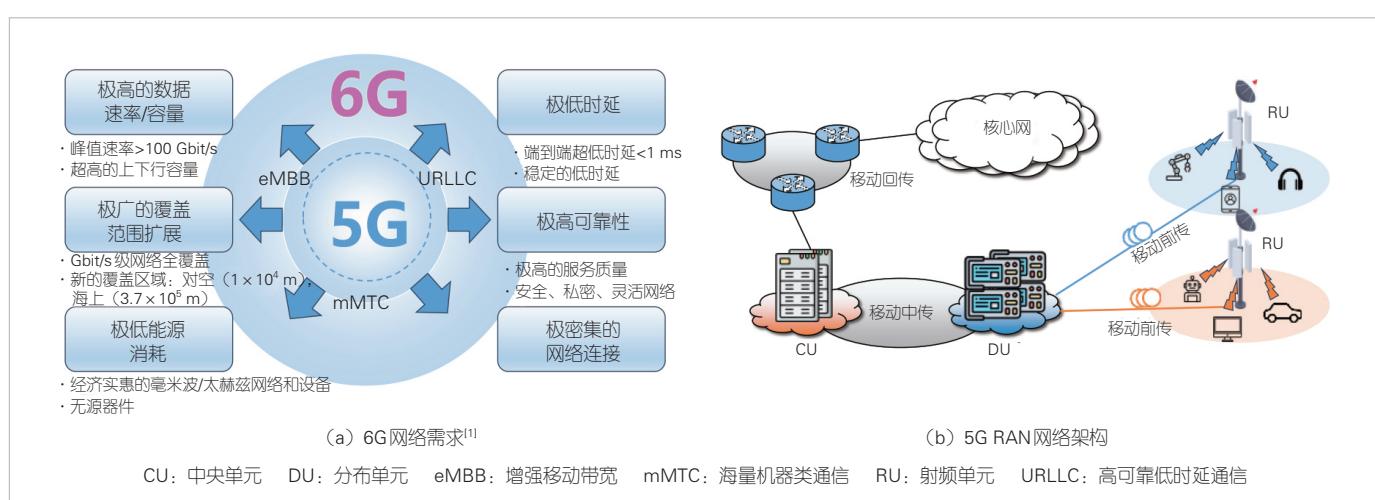
本文中，我们以未来前传网络的需求为导向，系统回顾和梳理了移动前传网络的需求、技术和实现架构等。此外，本文以光纤无线电融合接入为基础，重点针对数字前传技术、模拟前传技术和数模结合前传技术，围绕其网络架构、关键

技术和未来发展方向进行深入分析和探讨；进一步地，对未来前传网络关键技术进行展望，为未来移动前传网络的研究提供参考。

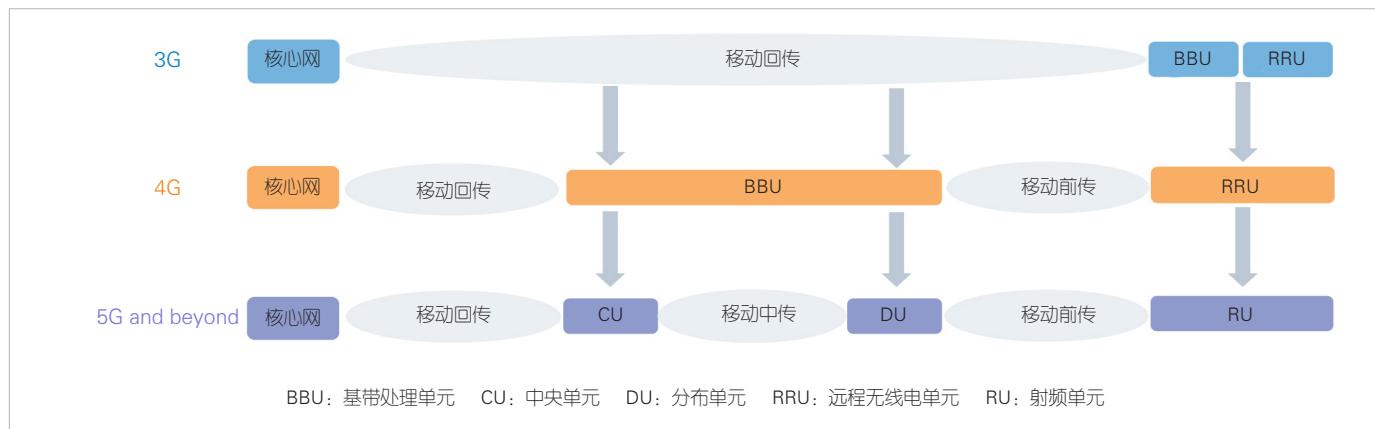
1 移动前传的发展

移动前传（MFH）是无线接入网功能划分的产物。5G乃至B5G的接入网架构如图1（b）所示。在接入网的发展中，随着接入网架构的变化，前传网络的功能定位也在随之改变。如图2所示，在3G接入网中，基带处理单元（BBU）和远程无线电单元（RRU）作为一个整体，通过回传（Backhaul）网络与核心网连接；4G时代，BBU与RRU被拆分开，它们之间的链接定义为前传（Fronthaul）网络，一般由光纤承载；5G时代后，接入网又被进一步拆分，BBU被拆分成了中央单元（CU）和分布单元（DU）。CU和DU之间形成了新的链路称为中传（Midhaul），而RRU也由于功能的增加演变成了射频单元（RU）。

为了实现更加有效的功能划分，第3代合作伙伴计划



▲图1 6G网络需求和5G RAN网络架构



▲图2 不同代际移动通信网络的前传网络网架构

(3GPP) 以网络层、数据链接层和物理层为基础提出了8种不同的功能划分选项，如图3所示。业内可根据实用需求，以不同功能划分为基础定义接入网。随着技术的发展和不同需求的增加，以功能划分为原则，形成了许多种不同结构的接入网，如分布式无线接入网、集中式无线接入网、云无线接入网和开放式无线接入网等。前传网络的功能定位也随之改变，形成了图2中3G、4G、5G乃至B5G的不同接入网架构。这种网络架构的演进既是网络功能的细化和功能分解的需要，也是系统对带宽、时延和处理能力的进一步优化。展望未来通信网络，前传网络将在5G的基础上，进一步提升传输带宽，降低时延，并展现高度的灵活性。

在前传网络之间的传输、连接和控制标准规范方面，业内根据不同实际需求在不同的功能区间制定了多种前传标准协议接口。4G前传标准接口是通用公共无线电接口(CPRI)，由原始设备制造商联盟于2003年制定。CPRI通常应用于选项8，使用同步数据传输协议，其比特率取决于天线数量。对于一个多单元多天线系统，CPRI的比特率可由式(1)表示：

$$B_{\text{CPRI}} = S \cdot A \cdot f_s \cdot b_s \cdot 2 \cdot \left(\frac{16}{15}\right) \cdot LC \quad (1)$$

其中，S为单元数，A为每个单元中的天线数量， f_s 表示采样率， b_s 表示每个样本的比特数（一般取15或8），参数2表示I和Q两路处理，16/15表示多余的开销信息，LC为约束编码（一般取10/8或66/64，取决于CPRI网络比特率选项）。

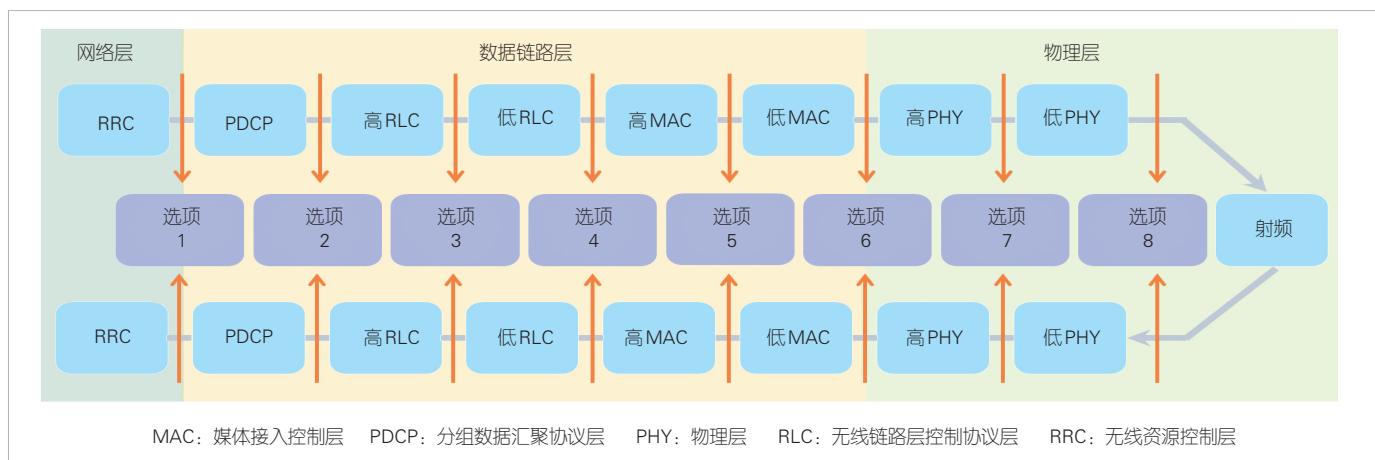
由式(1)可知，CPRI协议下的传输速率不会根据实际网络负载的变化而进行调整，这是产生CPRI低利用率的主要原因。4G前传网络主要采用CPRI协议，但由于5G网络速率需求大幅提升、天线数量成倍增加，CPRI难以满足5G高容量需求。4G网络中载波带宽一般为20 MHz，快速傅里叶

逆变换(IFFT)点数为2 048，子载波间隔为15 kHz，采样频率为30.72 MHz ($2 048 \times 15 \text{ kHz}$)，采样位宽为30 bit (I/Q各15 bit)。根据式(1)可以算出每单元单天线下所需的CPRI速率为1.228 8 Gbit/s。如果采用 8×8 多输入多输出(MIMO)，单条线路的CPRI速率需求就为9.830 4 Gbit/s ($8 \times 1.228 8 \text{ Gbit/s}$)。5G新空口定义载波带宽为100 MHz，IFFT点数为4 096，子载波间隔为30 kHz，采样频率为122.88 MHz ($4 096 \times 30 \text{ kHz}$)，采样位宽为30 bit (I/Q各15 bit)，可以算出每单元单天线下所需的CPRI速率为4.915 2 Gbit/s，而5G采用64TRx MIMO时，最终单条线路的CPRI速率需求为314.57 Gbit/s ($=64 \times 4.915 2 \text{ Gbit/s}$)。5G的CPRI速率需求较4G增长了几十倍。为此，3GPP TR 38.801、电气和电子工程师协会(IEEE)和国际电信联盟标准化部门ITU-T等组织制定了新的解决方案，例如将BBU的功能进行拆分，让部分功能下沉至RRU中，以降低接口之间的传输速率需求。

随着功能结构的拆分，新的接口协议也由此诞生。增强型通用公共无线电接口(eCPRI)，是用于连接无线基站DU与RU的前传接口协议，以物理层功能内部拆分为基础。相比于CPRI，eCPRI在选项7内新增了选项7-1、7-2a、7-2和7-3，目的就是将调制、映射、预编码以及循环前缀添加等数字信号处理(DSP)功能下沉到RU中，以此来减轻前传传输速率的压力，大幅提升前传网络的高效性和灵活性。新一代前传光网络正在如上的功能拆分的基础上，进一步创新前传光网络架构，发展物理层前传通信技术，满足未来前传的带宽、效率和灵活性需求。

2 相关工作

随着前传网络更深层次的覆盖与应用，业界对未来前传网络的发展趋势展开了一系列的研究。例如，文献[11]预测



▲图3 新一代移动前传网络的不同功能划分选项

未来前传网络会向着更加开放智能的方向发展，认为下一代移动通信可能会出现以用户为中心的网络架构，主要由人工智能和机器学习算法驱动，其网络端点可以根据用户操作习惯做出自主的网络决策。这种网络架构可以极大地减少用户端点与基站之间的通信开销。文献[12]对未来接入网以及前传网络发展趋势提出了3点看法：1) 未来接入网应该进一步虚拟化并且注重优化物理层管理与维护，以缓解多用户链路的端到端网络切片问题；2) 未来前传网络应更具有针对性，根据具体服务项目的需求分配专门的前传链路负责，以更好地满足用户的需求；3) 未来前传网络需要实现不同接口间的互操作性，这对搭建低运营成本、高传输效率的前传网络架构至关重要。文献[13]预测未来前传网络开放操作维护管理（OAM）接口将成为趋势，这能够做到对网络资源分配的实时监控，以实现网络资源的智能动态重构，从而满足网络切片之间共享资源的需求。在无线前传网络方面，文献[14]仿真研究了80 GHz高频段毫米波应用于无线前传的可能性，认为在经济性和实用性的考量下，未来无线前传网络可以部分替代光纤前传网络，提供稳定高速的传输链路。

下面本文将重点从新型前传光网络实现架构和光无线融合的新型前传光信号调制与传输技术两个方面，分析前传光网络的发展。

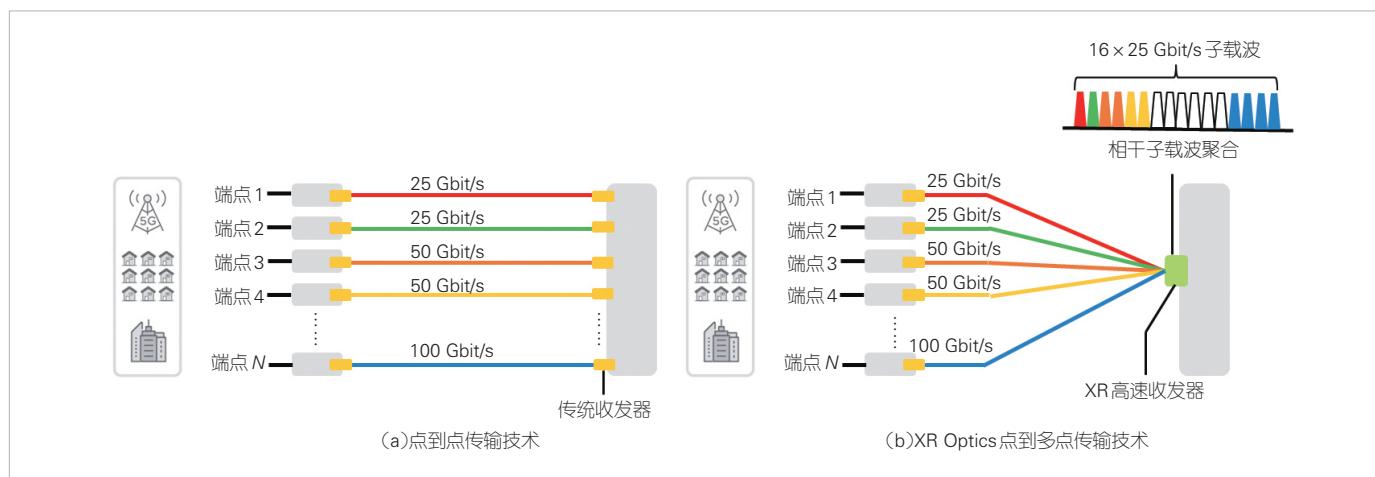
3 新型前传光网络实现架构

3.1 载波聚合的点到多点相干技术

前传光网络通常应用点对点（PtP）连接，如图4所示。点对点接口部署在许多网段中，市场规模巨大且有不断增长的趋势。在2020年的统计中，点对点接口占光组件总量的40%以上^[15]，可预计未来带宽25 GHz以上前传光接口

（CPRI或eCPRI）将占据主要市场。目前的5G前传通信大多基于点对点强度调制/直接检测（IM/DD）系统，采用专用光纤链路或波分复用（WDM）链路进行传输^[12]，每个RU通过单向或双向光纤直接与DU相连。该系统仅适用于光纤资源丰富的场景。考虑到基站建设愈加密集的趋势，部署大规模光纤的成本高昂，并且伴随B5G大规模MIMO的进一步部署，前传网络所需速率将变得前所未有的高。面对不断提升的速率要求，ITU-T目前已经提出了10 Gbit/s和25 Gbit/s的标准，50 Gbit/s的新标准正在准备中，IEEE的802.3标准也在不断完善中^[16]。但点对点IM/DD系统仍面临未来前传演进的巨大挑战。首先，IM/DD系统将遇到超过100 Gbit/s级别的容量瓶颈。随着波特率的提高，色散将从根本上限制IM/DD系统的性能。具体来说，IM/DD PAM4信号可传输的距离与波特率的平方成反比关系^[17]。对于一个波分复用IM/DD系统（每个通道200 Gbit/s）来说，其在o波段的传输范围被限制在大约2.5 km内（< 1 dB损失）^[18]。其次，灵活性是第二大限制。如图4（a）所示，在点对点架构中，每个终端会根据流量需求的峰值占用相应的资源，每个端点对应的链路都需要以相同的速度运行，导致整个网络成为一个极其低效的传输结构。实际过程中的实时流量可能会有所不同，点对点架构难以实现资源的有效分配而造成不必要的资源浪费。

面对这一问题，业内提出了基于相干载波聚合的点到多点相干技术。如Infinera在ECOC 2019上推出了XR Optics点到多点相干传输技术。如图4（b）所示，XR Optics利用数字信号处理技术将给定波长光谱的传输和接收细分为一系列称为相干数字子载波的频率通道，这些数字子载波可以独立地分配到不同的端点。这一技术实现了业界首个可扩展的点到多点和低速高速可相互转换的光模块连接。单个400G XR Optics集线器模块可生成16 × 25 Gbit/s数字子载波。一



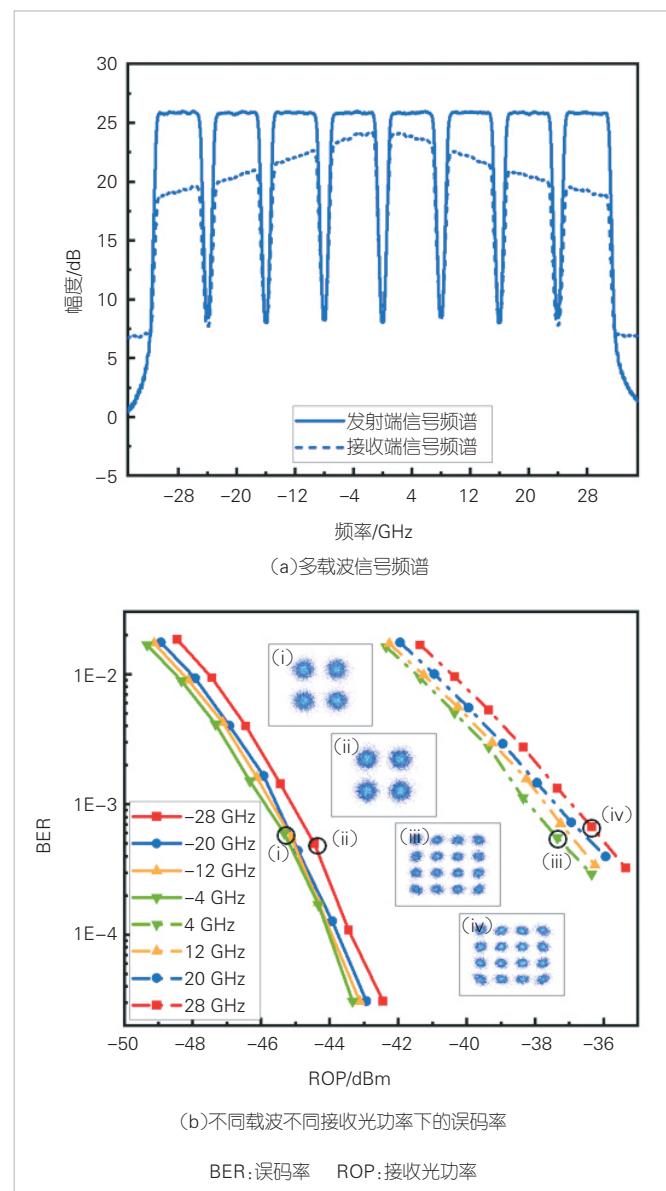
▲图4 点到点、点到多点传输技术

个或多个数字子载波可以组合并分配到特定的目的地，以提供需要的带宽。XR Optics 打破了传统点对点光学传输解决方案的局限性，引入了一种新型可插拔和支持软件的架构，从根本上降低了部署和运营光网络的成本。另外，XR Optics 器件提高了部署灵活性，相同的相干可插拔器件可以通过软件配置在点对点或点对多点系统中运行^[19]。这种方式支持灵活速率分配、灵活调制格式切换和点到多点的连接。如图 5 (a) 所示，我们仿真了基于相干数字子载波的载波聚合点到多点相干接入方案。这里，我们展示了 8 个数字子载波的相干收发，其中前 4 个子载波承载正交相移键控 (QPSK) 信号，后 4 个子载波承载 16-QAM 信号。图 5 (b) 表明，它们在较低的接收光功率情况下也能满足 $1E-2$ 的误码率门限要求。这证明了相干子载波点到多点系统实际应用的可行性。

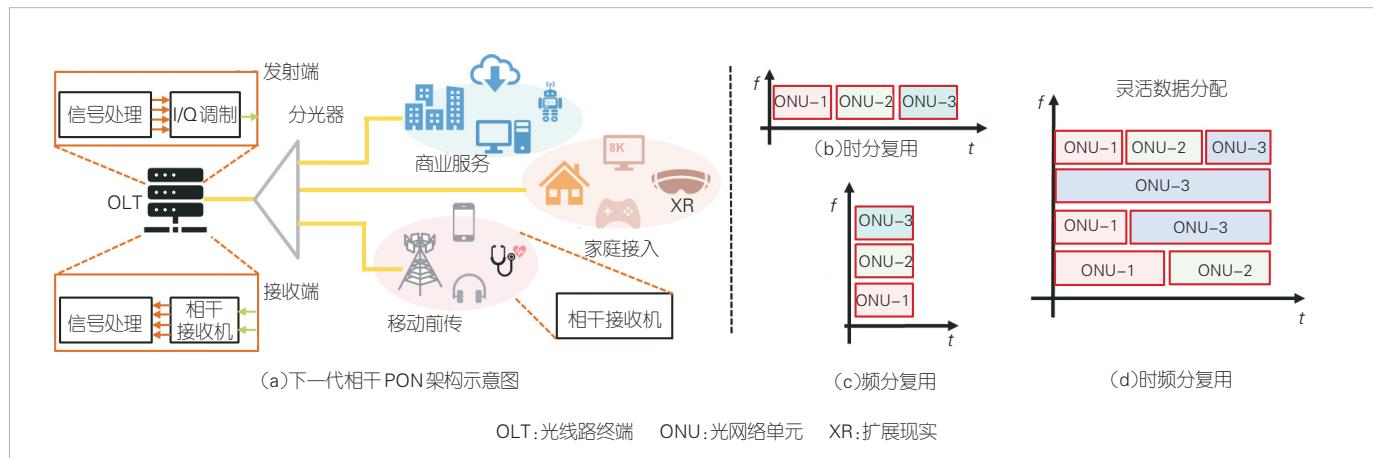
3.2 下一代相干无源光网络技术

未来移动通信、云计算、高清视频等高端业务的发展对前传光网络的传输速率提出了更高的需求。无源光网络 (PON) 技术是一种经济的小型密集部署的候选技术，大量应用在前传光网络中。在过去的几年中，IEEE 和 ITU-T 积极推进高速 PON 的标准化进程，可以推测未来 PON 系统将需要满足 100 Gbit/s、200 Gbit/s 甚至更高的传输速率需求。目前，IM/DD 技术是商用 PON 系统中的常用技术，部署简单，成本低；但是，由于灵敏度和功率预算受限，其很难满足未来高速传输的需求。相比而言，相干技术是适合下一代大容量 PON 系统的一项很有潜力的候选技术，其灵敏度高，可实现多维复用，结合先进数字信号处理 (DSP) 技术可以实现高速数据传输^[20-22]。

图 6 (a) 为下一代相干 PON 架构的示意图。光线路终



▲图 5 基于相干数字子载波的载波聚合点到多点仿真结果



▲图 6 下一代相干无源光网络架构及其复用方式

端的信号经过无源节点分配给不同的用户，提供商业服务、住宅接入或者移动前传等服务。在光线路终端的发射端，原始信号经过发端信号处理后被调制在I/Q调制器上，可实现信号的多维调制；在光线路终端的接收端，信号在集成相干接收机中进行探测，本地振荡信号的引入可以提高系统的接收灵敏度，探测得到的信号进行后续的信号处理。相应地，在光用户单元中也部署了相干收发机，从而构成了相干PON系统。在相干PON系统中常用的3种复用方式包括：时分复用（TDM）、频分复用（FDM）和时频分复用（TFDM）技术，如图6（b）–（d）所示。在TDM技术中，信号通过不同的时隙传输给不同的用户，可提高大带宽接入；在FDM技术中，信号被调制在不同的频率上，可提供低时延服务；TFDM结合两者的特点，信号可以在时间和带宽上进行灵活分配，提供更大的自由度。

目前，高速相干PON系统的研究已经取得了很多的进展。文献[23–24]基于相干探测和TDM技术实现了单波100G信号传输。在相干PON系统中，突发接收是一个重要的挑战，其中包括突发放大和突发信号处理。文献[25]研究了基于本振功率调整的突发放大技术，该技术可以有效提升突发相干PON系统的接收动态范围，实现了单波200G信号传输。文献[26]提出了一种基于有效前导的突发DSP技术，可以实现算法的快速收敛，并且在单波100G的TDM-PON系统中进行了验证。同时，相干PON系统的灵活性是一个值得关注的特性。用户端相对于光线路终端的距离是不一样的，导致用户端的接收光功率不一样。距离近的用户具有更高的功率来满足更高速率的传输，因此通过灵活的编码整形调制技术，可以进一步提高系统传输的容量，实现速率的灵活分配。文献[27]首次提出并演示了一种新颖的三维柔性相干PON（3D FLCS-CPON）。该技术具有在时域、频域和功率域内的资源分配能力，在20 km光纤上实现了 $250 \text{ Gbit} \cdot \text{s}^{-1} \cdot \lambda^{-1}$ 的峰值数据速率^[27]。文献[28]首次提出了基于概率整形技术的新一代灵活相干接入网框架，并结合本振光（LO）动态调控技术

实现了最高300G的高速光传输和 $7.104 \text{ dB} \cdot \text{Gbit} \cdot \text{s}^{-1}$ 的动态范围与净速率积（DRNRP）；基于灵活的调制格式切换，可以实现最高300 Gbit/s的速率传输^[29]。TFDM技术的研究也取得了很多的进展，例如：文献[30–31]利用数字子载波复用技术，在4个子通道中实现了100 Gbit/s的数据传输；文献[32]基于简化相干技术实现了实时TFDM-PON系统验证，上下行支持100 Gbit/s和200 Gbit/s的峰值速率传输。可以看出，相干PON系统能够实现高传输速率。灵活的速率调节与带宽分配，是未来高速率、低成本与低时延移动前传网络的支撑。

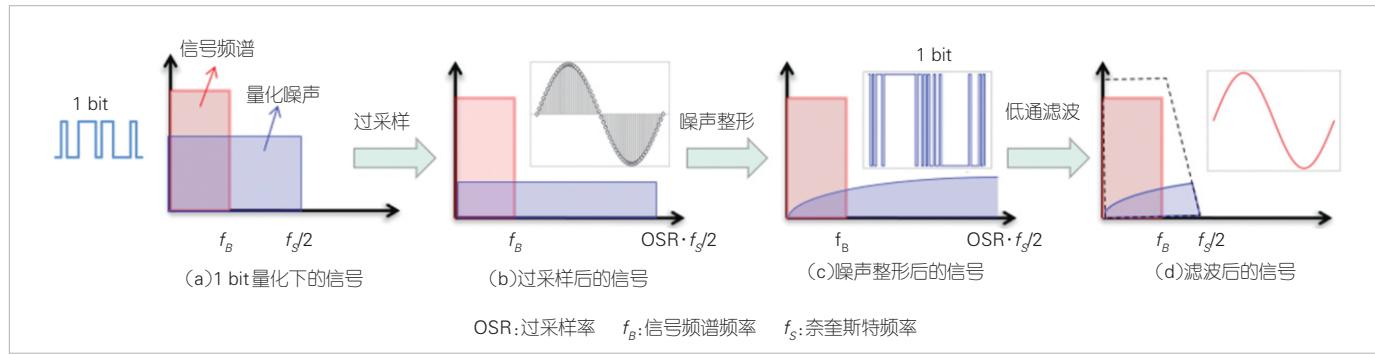
4 新型前传光信号调制与传输技术

相较于CPRI的数字前传接口，光载无线（RoF）融合传输技术具有带宽大、效率高和时延低等优点，在前传网络中具有广阔的应用前景。基于RoF技术，业内发展了多种新型前传光信号调制和传输技术。根据实现方式的不同，这些技术主要分为以下几类：数字前传技术、模拟前传技术和数模结合前传技术。

4.1 数字前传技术

数字前传技术拥有成熟、传输可靠、易于标准化等特点，CPRI、eCPRI等标准已经落地多年，拥有现成的设备。数字光纤无线电（D-RoF）目前大多部署在基于CPRI或eCPRI标准的前传链路上。数字信号传输受光纤色散、非线性等影响小，均衡手段丰富，美中不足的是数字信号面临带宽效率低、可扩展性有限和接收端设备复杂度高等问题。近年来，业界对数字信号前传的研究主要集中于提高信号信噪比，这不仅可满足高阶QAM需求从而增加每符号携带的比特数，还有利于实现高波特率信号的传输。

Delta-Sigma调制技术在数字前传领域的研究引人注目^[33]。如图7所示，Delta-Sigma调制技术可以将模拟信号量化成1 bit或者2 bit数字信号，该技术通过过采样和噪声整



▲图7 Delta-Sigma调制技术对信号的处理过程

形操作有效地降低量化噪声，并且通过模拟滤波器就可恢复出模拟信号，不需要额外的数模转换器（DAC）设备。在最新的研究中，上海大学提出将Delta-Sigma调制与自零相干接收机结合，可以做到2.5 Gbaud 4194304-QAM的相干光纤系统传输^[33]。华中科技大学提出将分层调制技术与Delta-Sigma技术相结合，在不占用额外频谱资源的情况下使访问容量增大一倍，并且可以根据需求灵活分配接收端设备的信噪比^[34]。同时，他们还提出了两层Delta-Sigma调制的方法，相较于一般Delta-Sigma调制在IM/DD系统中有15.9 dB的信噪比提升^[35]。另外，复旦大学提出了概率整形（PS）与Delta-Sigma结合技术，实现了28.1 Gbit/s PS-131072-QAM的信号传输，进一步扩展了Delta-Sigma调制技术的适用性^[36]。从经济性方面考虑，在发射端，传统模数转换器（ADC）需要使用模拟低通滤波器来减少高于奈奎斯特频率的成分，这些滤波器需要具备严格性能要求，难以设计和制造，并且成本高。Delta-Sigma ADC采用过采样的方法，可以减轻对模拟抗混叠滤波器的要求。过采样使得输入信号的频谱在频域中被分离开来，允许使用更宽带宽、更缓慢截止特性的滤波器，降低了滤波器设计的难度。在接收端，Delta-Sigma信号只需要通过模拟滤波器就可以完成数模转换，进一步简化了RU结构。

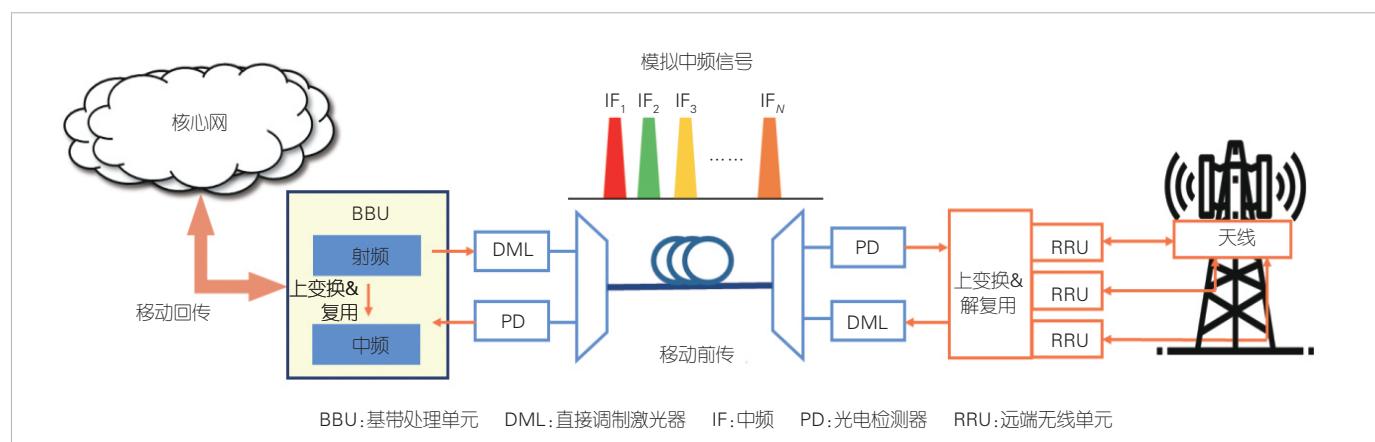
4.2 模拟前传技术

模拟前传技术直接将无线信号调制到光载波上，较于数字前传技术能够极大地提高信号频谱效率，并有望实现超低时延的前传接入。另外，模拟前传技术还拥有基础设施简单和成本低的特点，可以在很大程度上简化RU设备。这些特点对部署大量密集的小型单元天线站点至关重要。然而，模拟信号过程中容易出现信号失真，从而影响前传网络的性

能。但该方案仍是移动前传领域的有力竞争者，近年来得到了广泛的研究。

在信号处理过程中使用数字信号处理手段可以实现模拟信号的均衡，但常规的数字信号处理手段难以取得较好的效果，因此，关于模拟前传技术的研究大多集中于模拟信号的均衡方法。最新的研究中，业内分别尝试了使用神经网络数字预失真算法^[37]和支持向量机（SVM）^[38]来均衡模拟信号。奥尔堡大学提出了一种基于机器学习（ML）的数字预失真（DPD）解决方案^[39]，提出的支持向量回归（SVR）方法可以有效地缓解非线性，提高模拟信号信噪比，与未应用DPD方法相比，SVR-DPD方案的矢量幅度误差（EVM）降低了6.9%。与基于机器学习方案相比，阿尔斯特大学提出了一种新的基于幅度选择反射（MSA）的DPD方法^[40]，在规范分段线性（CPWL）模型的基础上，通过引入MSA函数，减少乘法运算的次数，降低模型的复杂性。除了均衡手段外，还有人提出了将IFFT前的频域采样技术和IFFT后的时域采样技术相结合的模拟子载波复用技术，该技术可在一定采样率内灵活地复用和解复用^[41]。

此外，A-RoF波分复用技术中频光纤无线电（IFoF）也是A-RoF领域的研究热点之一，如图8所示。中频是指略低于射频的载波频率。比起RoF，IFoF有着更灵活的带宽分配优势，产生的较低频的信号光纤色散影响较小。但这一技术需要在RU中加入射频中频的变频器，这会增加RU的复杂度。关于IFoF技术，早在2017年，文献[42]就成功验证了在基于IFoF的移动前传技术下实现5G移动通信传输的可能性，例如：在进行4K视频流信号的实时传输时，每位用户的峰值数据速率高达1.5 Gbit/s，这符合IMT-2020关于5G愿景的要求^[42]。日本凯迪迪爱研究所（KDDI）使用并行强度调制/相位调制发射机在20 km单模光纤上传输了14×1.2 GHz



▲图8 模拟中频光纤无线电系统

OFDM信号，实现了CPRI等效数据速率1.032 Tbit/s^[43]。并行IM/PM发射机在面对不同的调制信号频率时，会根据光纤色散零频率来选择信号应该进行IM调制还是PM调制，从而有效避免光纤色散带来的影响，这为模拟信号的光纤传输提供了很好的解决方案。

4.3 数模结合前传技术

为了结合频谱高效的A-RoF和高保真D-RoF的优点，业界也对数模结合RoF（DA-RoF）进行了研究。2021年，Futurewei Technologies提出了一种类似概率整形的数模信号结合方案^[8]，如图9所示。该算法通过将信号与数字放大倍数a相乘后四舍五入将信号量化，之后用原信号与量化后的信号相减得到的模拟信号，模拟信号可以与模拟放大倍数b相乘，之后两段信号时分复用以作为发射信号。这种方法通过补偿量化噪声来实现较高信噪比。对于数字信号部分，数字放大倍数决定了量化信号的阶数。阶数越高，量化噪声越小，但相应受到的信道噪声就越大。对于模拟信号部分，适当放大信号幅值有利于增强模拟信号部分的抗噪能力，但如果幅值过大，信号就会受到峰值平均功率比过大的影响，从而影响信号质量。对此，我们测试了在IM/DD BtB系统下传输20 Gbaud 1024-QAM OFDM信号时，不同数字放大倍数和模拟放大倍数情况下的信噪比。结果表明，信号在数字放大倍数a=4.5和模拟放大倍数b=1.5时信噪比最高可达35.5 dB。上海交通大学在这一算法的基础上，通过对信号多次四舍五入数字化操作，对3.2 GHz 1048576-QAM的高阶调制信号进行20 km单模光纤传输，使接收信号具有高达70.8 dB的信

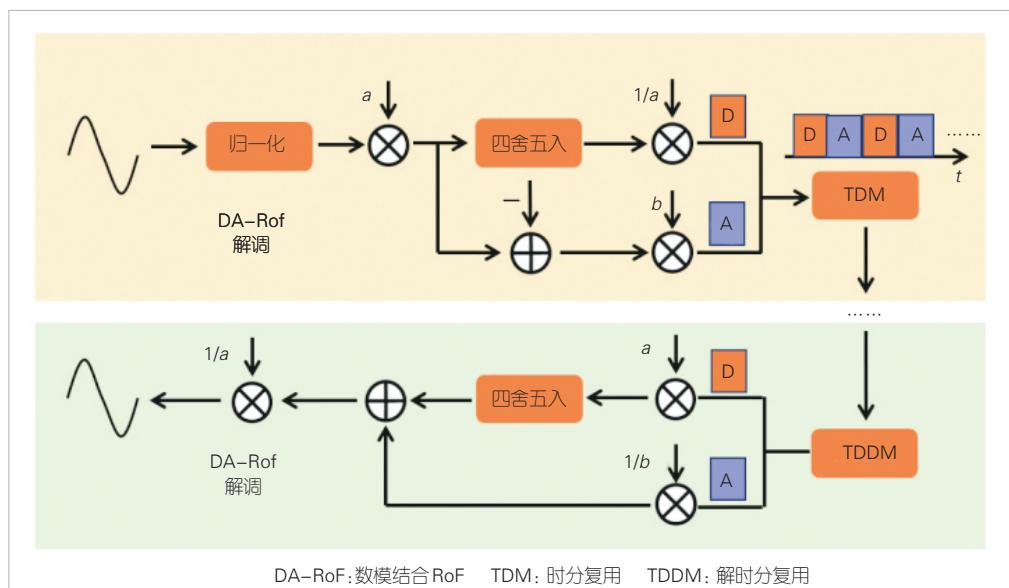
噪比^[44]。在相干检测中，传统频偏和相偏估计算法只针对有限个星座点信号进行跟踪，难以对模拟部分信号进行恢复。为解决这一问题，清华大学和上海交通大学利用光频梳谱线拥有间隔相等且稳定的特点，仅通过两条或三条导频就可以异地再生一组相同的光频梳，同时接收端将克隆的光频梳作为本振光，在WDM的同时有效地解决DA-RoF模拟部分的频偏、相偏问题。该方法同时支持14.1 Tbit/s公共无线电接口等效数据速率和1024-QAM，有助于实现基于光纤和自由空间的光前传，从而实现全频段的相干光接入网^[45]。在工程可行性方面，这种算法在调制解调每个符号时只需要线性地增加22次数字运算，就可以获得指数级信噪比的增长。每增加一倍带宽就可以获得接近10 dB的信噪比增益。该算法在未来低功耗、低时延网络中具有巨大的应用价值。

5 展望

尽管移动前传光网络在网络架构和具体信号调制与传输方面取得了一定的进展，但未来前传网络还存在一定的挑战。毫米波、太赫兹等新频谱通信的发展，深刻影响了前传网络的具体形式。此外，超低时延和超高灵活性等新需求也将促进未来前传网络的进一步发展。本文中，我们将从以下几个方面展望未来前传光网络：

1) 新型毫米波、太赫兹光无线基带前传与无缝融合技术。下一代移动通信的宏大发展愿景，离不开通信载体特别是物理层的深刻变革。针对6G超宽带、超低时延和极致覆盖的需求，新的频谱资源特别是太赫兹频段(0.1~10 THz)的通信技术，成为最受关注的6G核心技术之一，引起了全球业界和学术研究机构的高度关注。

太赫兹波段介于光和微波之间，以其丰富的频谱资源和独有特性，获得了ITU的大力支持，成为极具潜力的6G关键候选频谱技术。太赫兹通信具有极为丰富的应用场景，是未来实现超高速移动无线通信的关键技术，具有重要的战略和经济价值，成为全球科技竞争的制高点。然而，太赫兹波段频率高、带宽大，对相关器件的带宽、频率和工作效率方面提出了重大的挑战。此外，下一代移动通信网络



▲图9 数模信号结合方案

对整体功耗的要求进一步加强，传统射频微波电子器件难以满足要求。纯电链路的太赫兹通信系统存在相噪恶化和杂波干扰的问题。光生太赫兹及光子太赫兹混频融合的系统不仅能有效克服这一问题，还具有带宽大、上下变频信号纯净和调控灵活的优点。因此，一种集成光纤传输，光子毫米波太赫兹产生、调制和传输的光无线融合架构具有重要的优势。如图10所示，光子毫米波、光子太赫兹通信系统可以作为基于光纤接入网的无线接入中继和扩展，为用户提供无缝的“最后一公里”连接。CU实现下行信号的光调制和上行信号的相干探测，利用光纤网络实现基带信号前传的上下行通信。在远端的基站实现信号的上下变频。光子辅助的毫米波太赫兹信号产生和探测可通过宽带光电转化和电光转化完成。这种光无线无缝融合的架构能实现宽带光信号的基带前传功能，具有高度集成、无缝融合和高度灵活的特点。

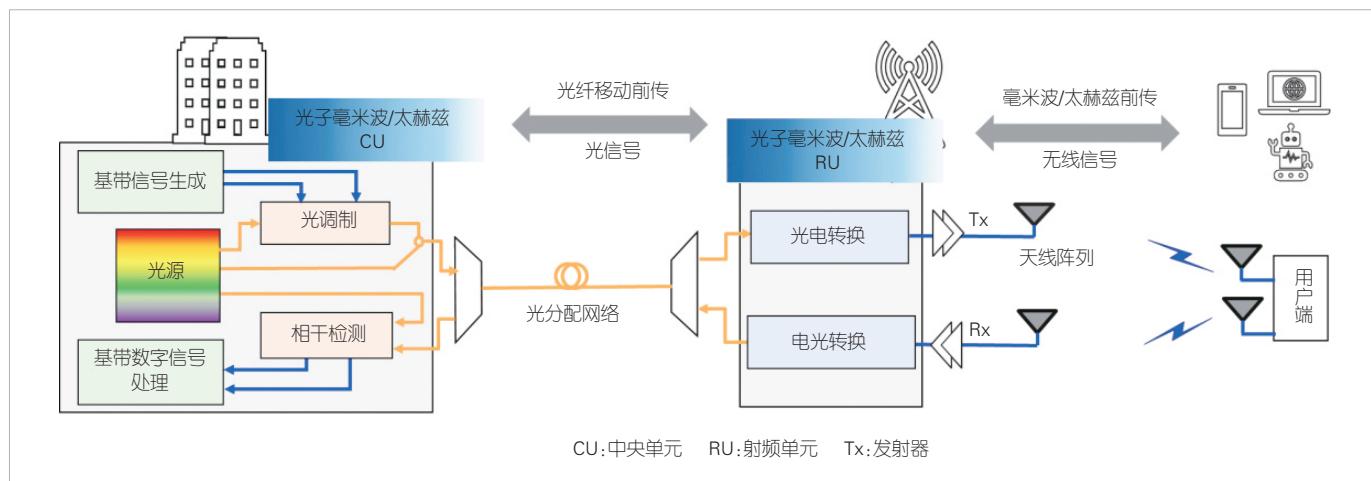
2) 超低时延、超高速和超高灵活性前传网络技术。未来6G网络将实现真实物理世界和虚拟数字世界的深度融合，面对沉浸式云扩展现实(XR)、全息通信、感官互联等新技术，前传光网络将承载比今天多数十倍乃至百倍的数据容量。对于实现低时延高速前传网络，一方面，少DSP甚至无DSP前传通信技术对降低时延具有重要的意义；另一方面，先进的DSP算法可以进一步提高信号的信噪比，实现超高阶信号调制，从而提高传输速率。然而，过于复杂的DSP手段必然会导致时延的上升。性能和时延之间的权衡也是未来前传网络面临的问题之一。对于实现超高灵活性网络，一方面，现今不同设备商的光通信设备仍无法做到统一的接口，运营商只能采用单一设备商的产品来部署光网络，因此未来前传光网络需要灵活、可互操作、即插即用的前传光器件，

以降低成本、提高性能和灵敏度；另一方面，未来前传光网络需要高效精准的资源分配机制，这是因为当前固定带宽分配的前传网络分配机制容易带来资源浪费、成本提高以及延迟过高等诸多问题^[46]。为了解决这一问题，可进一步研究无线接入网的虚拟化，将深度学习、人工智能等先进算法用于资源分配，建立以流量为导向的频谱资源分配机制；此外，也可以调整前传网络架构，建立基站与RU间的灵活映射关系，同时开启RU之间的数据共享，开发协作多点(CoMP)的前传技术等。

3) 天地一体化异构前传网络。未来前传网络将进一步扩展网络覆盖的广度与深度。光纤网络、毫米波、太赫兹等陆基网络和以低轨卫星通信网络为代表的非陆基网络，将在未来走向融合。未来，随着卫星互联网、高速手机直连、天空地一体化的发展，天基低轨卫星通信网将成为地面有线和无线通信网络之外的新一代接入网，承载越来越多的无线互联网流量，与地面移动通信网络互为补充，共同构建覆盖全球天地一体化的无线接入网络。如何实现和应用天地一体化异构前传网络，成为一个重要的课题。一方面，多种网络形态对前传网络提出了重要的挑战，单一前传技术难以满足异构网络的不同物理传输链路需求；另一方面，天地一体化亟需高效的网络间协同技术，通过优化网络间的协同工作和切换策略，可以得到更高的数据容量和更低的传输时延，这对前传网络的融合性提出了更高的要求。

6 结束语

未来下一代移动通信高速率低时延的应用需求与日俱增，这对移动前传光网络提出巨大挑战。本文系统回顾了



▲图10 毫米波/太赫兹移动前传网络

前传光网络的发展，讨论了未来前传光网络的接入方式和未来前传光网络相干PON的应用需求，并以RoF技术为基础，调研了数字前传技术、模拟前传技术和数模结合前传技术3个方向的前传前沿技术。最后，我们给出了前传网络的未来发展方向。移动前传光网络将是下一代移动通信网络实现高速率、低时延通信的关键一环。稳定、高效、无缝接入的前传光网络将为下一代移动通信的成功落地奠定重要基础。

参考文献

- [1] NTT DOCOMO. 5G evolution and 6G [EB/OL]. [2023-07-23]. https://www.nttdocomo.ne.jp/english/binary/pdf/corporate/technology/whitepaper_6g/DOCOMO_6G_White_PaperEN_20200124.pdf
- [2] RANAWERA C, WONG E, NIRMALATHAS A, et al. 5G C-RAN with optical fronthaul: an analysis from a deployment perspective [J]. *Journal of lightwave technology*, 2018, 36(11): 2059–2068. DOI: 10.1109/JLT.2017.278282
- [3] WELCH D, NAPOLI A, BÄCK J, et al. Point-to-multipoint optical networks using coherent digital subcarriers [J]. *Journal of lightwave technology*, 2021, 39(16): 5232–5247. DOI: 10.1109/JLT.2021.3097163
- [4] ZHANG J W, JIA Z S. Coherent passive optical networks for 100 G/λ-and-beyond fiber access: recent progress and outlook [J]. *IEEE network*, 2022, 36(2): 116–123. DOI: 10.1109/MNET.005.2100604
- [5] LAGÉN S, GIUPPONI L, HANSSON A, et al. Modulation compression in next generation RAN: air interface and fronthaul trade-offs [J]. *IEEE communications magazine*, 2021, 59(1): 89–95. DOI: 10.1109/MCOM.001.2000453
- [6] XU M, LU F, WANG J, et al. Key technologies for next-generation digital RoF mobile fronthaul with statistical data compression and multiband modulation [J]. *Journal of lightwave technology*, 2017, 35(17): 3671–3679. DOI: 10.1109/JLT.2017.2715003
- [7] LIU S M, ALFADHLI Y M, SHEN S Y, et al. A novel ANN equalizer to mitigate nonlinear interference in analog-RoF mobile fronthaul [J]. *IEEE photonics technology letters*, 2018, 30(19): 1675–1678. DOI: 10.1109/LPT.2018.2865529
- [8] LIU X. Hybrid digital-analog radio-over-fiber (DA-RoF) modulation and demodulation achieving a SNR gain over analog RoF of >10 dB at halved spectral efficiency [EB/OL]. [2023-07-23]. <https://opg.optica.org/abstract.cfm?URI=OFC-2021-Tu5D.4>
- [9] HASABELNABY M A, SELMY H A I, DESSOUKY M I. Joint optimal transceiver placement and resource allocation schemes for redirected cooperative hybrid FSO/mmW 5G fronthaul networks [J]. *Journal of optical communications and networking*, 2018, 10(12): 975–990
- [10] ZHANG L, CHEN Z F, ZHANG H Q, et al. Hybrid fiber – THz fronthaul supporting up to 16384-QAM-OFDM with the delta-sigma modulation [J]. *Optics letters*, 2022, 47(17): 4307. DOI: 10.1364/ol.466080
- [11] RANAWERA C, KUA J, DIAS I, et al. 4G to 6G: disruptions and drivers for optical access [J]. *Journal of optical communications and networking*, 2022, 14(2): A143–A153. DOI: 10.1364/JOCN.440798
- [12] SALIOU F, CHANCLOU P, NETO L A, et al. Optical access network interfaces for 5G and beyond [J]. *Journal of optical communications and networking*, 2021, 13(8): D32–D42
- [13] LARRABEITI D, CONTRERAS L M, OTERO G, et al. Toward end-to-end latency management of 5G network slicing and fronthaul traffic (Invited paper) [J]. *Optical fiber technology*, 2023, 76: 103220. DOI: 10.1016/j.yofte.2022.103220
- [14] JIANG M L, CEZANNE J, SAMPATH A, et al. Wireless fronthaul for 5G and future radio access networks: challenges and enabling technologies [J]. *IEEE wireless communications*, 2022, 29(2): 108–114. DOI: 10.1109/JPHOT.2021.3135148
- [15] DAS S, RUFFINI M. PON virtualisation with EAST–WEST communications for low-latency converged multi-access edge computing (MEC) [EB/OL]. [2023-7-30]. <https://arxiv.org/abs/2004.06138.pdf>
- [16] SALIOU F, NETO L A, SIMON G, et al. 5G & optics in 2020 – where are we now? what did we learn? [EB/OL]. [2023-07-20]. <https://ieeexplore.ieee.org/document/9333337>
- [17] XIE C J, CHENG J C. Coherent optics for data center networks [C]// Proceedings of 2020 IEEE Photonics Society Summer Topicals Meeting Series (SUM). IEEE, 2020: 1–2. DOI: 10.1109/SUM48678.2020.9161052
- [18] KUMAR S, PAPEN G, SCHMIDTKE K, et al. Intra-data center interconnects, networking, and architectures [C]//Optical Fiber Telecommunications VII. Elsevier, 2020: 627–672
- [19] Infinera. Innovation details: XR optics [EB/OL]. [2023-07-23]. <https://www.infinera.com/innovation/xr-optics>
- [20] FARUK M S, LI X, NESSET D, et al. Coherent passive optical networks: why, when, and how [J]. *IEEE communications magazine*, 2021, 59(12): 112–117. DOI: 10.1109/MCOM.010.2100503
- [21] ZHANG J W, JIA Z S. Coherent passive optical networks for 100 G/λ-and-beyond fiber access: recent progress and outlook [J]. *IEEE network*, 2022, 36(2): 116–123. DOI: 10.1109/MNET.005.2100604
- [22] CAMPOS L A, JIA Z S, ZHANG H P, et al. Coherent optics for access from P2P to P2MP [J]. *Journal of optical communications and networking*, 2023, 15(3): A114–A123
- [23] ZHU Y X, YI L L, YANG B, et al. Comparative study of cost-effective coherent and direct detection schemes for 100 Gb/s/λ PON [J]. *Journal of optical communications and networking*, 2020, 12(9): D36. DOI: 10.1364/jocn.390911
- [24] KOMA R, FUJIWARA M, KANI J I, et al. Burst-mode digital signal processing that pre-calculates FIR filter coefficients for digital coherent PON upstream [J]. *Journal of optical communications and networking*, 2018, 10(5): 461–470
- [25] LI G Q, XING S Z, JIA J L, et al. Local oscillator power adjustment-based adaptive amplification for coherent TDM-PON with wide dynamic range [J]. *Journal of lightwave technology*, 2023, 41(4): 1240–1249. DOI: 10.1109/JLT.2022.3216763
- [26] ZHANG J W, JIA Z S, XU M, et al. Efficient preamble design and digital signal processing in upstream burst-mode detection of 100G TDM coherent-PON [J]. *Journal of optical communications and networking*, 2021, 13(2): A135–A143
- [27] SHEN W W, XING S Z, LI G Q, et al. Demonstration of beyond 100G three-dimensional flexible coherent PON in downstream with time, frequency and power resource allocation capability [C]//Proceedings of Optical Fiber Communication Conference (OFC) 2023. Optica Publishing Group, 2023: 1–3. DOI: 10.1364/ofc.2023.w1i.5
- [28] XING S Z, LI G Q, SUN A L, et al. Demonstration of PS-QAM based flexible coherent PON in burst-mode with 300G peak rate and ultra-wide dynamic range [J]. *Journal of lightwave technology*, 2023, 41(4): 1230–1239. DOI: 10.1109/JLT.2022.3208575
- [29] LI G, YAN A, XING S, et al. Pilot-aided continuous digital signal processing for multi-format flexible coherent TDM-PON in downstream [C]//2023 Optical Fiber Communications Conference and Exhibition (OFC). IEEE, 2023: 1–3
- [30] ZHANG J, JIA Z, ZHANG H, et al. Rate-flexible single-wavelength TFDM 100G coherent PON based on digital subcarrier multiplexing technology [EB/OL]. [2023-07-23]. <https://opg.optica.org/abstract.cfm?URI=OFC-2020-W1E.5>
- [31] XU M, JIA Z S, ZHANG H P, et al. Intelligent burst receiving control in 100G coherent PON with 4 × 25G TFDM upstream transmission [C]// Proceedings of 2022 Optical Fiber Communications Conference and Exhibition (OFC). IEEE, 2022: 1–3
- [32] WANG H D, ZHOU J, XING Z P, et al. Fast-convergence digital signal processing for coherent PON using digital SCM [J]. *Journal of lightwave technology*, 2023, 41(14): 4635–4643. DOI: 10.1109/JLT.2023.3243828
- [33] LUO S Y, LI Z X, FAN C, et al. Digital mobile fronthaul based on delta-sigma modulation employing a simple self-coherent receiver [J]. *Optics express*, 2022, 30(17): 30684. DOI: 10.1364/oe.459976
- [34] ZOU Y, ZHONG L S, ZHANG S M, et al. A hierarchical modulation enabled SNR allocable delta-sigma digital mobile fronthaul system [J]. *IEEE photonics journal*, 2021, 14(1): 1–6. DOI: 10.1109/JPHOT.2021.3135148

- [35] ZHONG L S, ZOU Y, ZHANG S M, et al. Experimental demonstration of delta-sigma modulation supported 65536-QAM OFDM transmission for fronthaul/WiFi applications [C]//Proceedings of 2021 European Conference on Optical Communication (ECOC). IEEE, 2021: 1–3. DOI: 10.1109/ECOC52684.2021.9606072
- [36] ZHAO L, XU S C, WANG M X, et al. Probabilistic shaping-based delta sigma modulation [J]. Optics letters, 2023, 48(6): 1450–1453
- [37] HADI M U, MURTAZA G. Enhancing distributed feedback-standard single mode fiber-radio over fiber links performance by neural network digital predistortion [J]. Microwave and optical technology letters, 2021, 63(5): 1558–1565. DOI: 10.1002/mop.32774
- [38] HADI M U. Mitigation of nonlinearities in analog radio over fiber links using machine learning approach [J]. ICT express, 2021, 7(2): 253–258. DOI: 10.1016/j.icte.2020.11.002
- [39] HADI M U, BASIT A. Machine learning for performance enhancement in fronthaul links for IOT applications [C]//Proceedings of 2021 International Conference on Digital Futures and Transformative Technologies (ICoDT2). IEEE, 2021: 1–5. DOI: 10.1109/ICoDT252288.2021.9441542
- [40] HADI M U, SOIN N, KAUSAR S. Enhancing 5Gmulti-band long haul optical fronthaul links performance by magnitude-selective affine digital predistortion method [J]. Microwave and optical technology letters, 2022, 64(4): 827–834. DOI: 10.1002/mop.33169
- [41] ASSIMAKOPoulos P, NOOR S, WANG M Q, et al. Flexible and efficient DSP-assisted subcarrier multiplexing for an analog mobile fronthaul [J]. IEEE photonics technology letters, 2021, 33(5): 267–270. DOI: 10.1109/LPT.2021.3056511
- [42] SUNG M, CHO S H, KIM J, et al. Demonstration of IFoF-based mobile fronthaul in 5G prototype with 28-GHz millimeter wave [J]. Journal of lightwave technology, 2018, 36(2): 601–609. DOI: 10.1109/JLT.2017.2763156
- [43] ISHIMURA S, BEKKALI A, TANAKA K, et al. 1.032-Tb/s CPRI-equivalent rate IF-over-fiber transmission using a parallel IM/PM transmitter for high-capacity mobile fronthaul links [J]. Journal of lightwave technology, 2018, 36(8): 1478–1484. DOI: 10.1109/JLT.2017.2787151
- [44] ZHU Y, ZHUGE Q, HU W. Efficient SNR scaling at >10 dB per extra bandwidth using cascaded hybrid digital-analog radio-over-fiber for fronthaul [C]//2022 27th OptoElectronics and Communications Conference (OECC) and 2022 International Conference on Photonics in Switching and Computing (PSC). IEEE, 2022: 1–3
- [45] ZHANG C, ZHU Y, HE B, et al. Clone-comb-enabled high-capacity digital-analogue fronthaul with high-order modulation formats [EB/OL]. (2023-08-31) [2023-09-05]. <https://www.nature.com/articles/s41566-023-01273-2>
- [46] ZHANG J W, JI Y F. 5G optical fronthaul: key issues, features and goals [J]. Scientia sinica informationis, 2017, 47(10): 1435–1442. DOI: 10.1360/n112017-00136

作者简介



赵俊皓, 复旦大学在读硕士研究生; 主要从事6G高速光接入网和光子无线融合的研究工作, 包括超高速光通信算法与系统、智能光子信号处理与机器学习智能光网络等。



张俊文, 复旦大学通信科学与工程系研究员、马可尼青年学者、国家特聘青年专家、上海市特聘专家、OFC 2022 N4 SC主席、2018年美国光学学会数字系统光学技术组主席、2019年IEEE SUM会议接入网分会主席, 连续多年担任OFC、SPIE-PW等国际大会技术委员, 并担任《IEEE Photonics Journal》《Frontiers in Communications and Networks》副编辑; 长期从事高速光传输、光无线融合和光接入网的研究, 在下一代高速光传输、光接入以及无线融合方面取得突出成就; 先后获得马可尼青年学者奖、第三届全国光学工程学会优秀博士论文提名奖、IEEE光子学会博士奖、王大珩光学奖等奖项; 发表论文200余篇, 出版英文专著1部, 申请美国专利/PCT专利22项、中国专利1项。



迟楠, 复旦大学教授、信息科学与工程学院院长, 教育部2006年度新世纪优秀人才, 2007年度湖北省杰出青年基金获得者, 2008年上海市曙光学者, 日本大川情报基金获得者, 2009年上海市浦江人才, 2019年国家杰出青年科学基金获得者, 美国光学学会OSA Fellow, 中国电子学会女科学家俱乐部轮值主席, 中国光学学会纤维与集成光学专委会常委委员, 中国通信学会光通信专业委员会委员; 主要研究方向是高速调制光传输机制与技术、光与无线融合网络和光电子集成器件; 在可见光通信研究方面, 研制了中国首个高速实时可见光通信系统样机, 该样机在第15届国际工业博览会展出并获得创新奖。

基于机器学习的智能路由解释方法



Interpreting Machine Learning-Based Intelligent Routing Algorithms

孟子立/MENG Zili, 徐明伟/XU Mingwei

(清华大学, 中国 北京 100084)
(Tsinghua University, Beijing 100084, China)

DOI: 10.12142/ZTETJ.202305009

网络出版地址: <http://kns.cnki.net/kcms/detail/34.1228.TN.20231017.0948.002.html>

网络出版日期: 2023-10-17

收稿日期: 2023-08-02

摘要: 综述了基于机器学习的智能路由方法的进展，并提出了一种针对基于机器学习的智能路由技术的解释方法。该方法可以对神经网络等黑盒子技术的输出决策结果进行解释，支持几乎所有类型的智能路由算法。网络管理员可以利用该方法理解智能路由算法为何做出某些决策，并在此技术上进一步优化算法，排除故障，增强部署信心。

关键词: 智能路由；图神经网络；超图；可解释性

Abstract: The recent advances in machine-learning-based intelligent routing algorithms are reviewed and a new interpretation method for machine-learning-based intelligent routing algorithms is proposed. This method can explain output decision results of black-box technologies such as neural networks. Network operators can therefore utilize such an interpretation method to understand the logic behind it. Further optimizations of the algorithm, debugging, and enhancing the confidence of deployment can be explored.

Keywords: intelligent routing; graph neural network; hypergraph; interpretability

引用格式: 孟子立, 徐明伟. 基于机器学习的智能路由解释方法 [J]. 中兴通讯技术, 2023, 29(5): 56–60. DOI: 10.12142/ZTETJ.202305009

Citation: MENG Z L, XU M W. Interpreting machine learning-based intelligent routing algorithms [J]. ZTE technology journal, 2023, 29(5): 56–60. DOI: 10.12142/ZTETJ.202305009

路由是网络中的重要科学问题，是数据包从一端到另一端寻路所必需的组成部分。近年来，智能路由技术也得到业界越来越多的关注。智能路由技术智能的智能性主要体现在路由算法开始采用机器学习、深度学习等技术，并使用了图神经网络、强化学习等建模、训练方法来对路由算法进行优化。

然而，随着机器学习等复杂算法的引入，路由的决策过程逐渐变得黑盒化、不透明。因此，尽管网络管理员知道某神经网络的路由算法较优，但可能较难理解其背后的逻辑，从而不敢轻易相信。因此，基于机器学习的智能路由技术面临部署困难等问题。

1 基于机器学习的智能路由主要技术

一方面，人工智能技术在近年来发展十分迅猛；另一方

面，软件定义网络与可编程路由设备的研究甚至规模部署也为基于机器学习的复杂路由算法提供了部署的可能。本节将按照机器学习技术的分类，从监督学习和强化学习两种常用的机器学习方法对智能路由的主要技术进行介绍。

1.1 基于监督学习的技术

监督学习是一种预先将网络状态和与其对应的较优的路由策略标记出来的技术。该技术使得模型能够准确地完成输入到输出映射。目前监督学习的趋势是所采用的模型基本上是深度神经网络。近年来，业界提出了一些基于监督学习的路由方法，这些方法要么直接采用深度神经网络进行决策优化，要么引入一些深度学习的模块辅助现有算法进行决策。

在直接使用深度神经网络进行决策的解决方案中，常见的做法是将网络拓扑（网络信息）与数据包序列号、流量需求矩阵、链路利用率、延迟、吞吐量、流量特征等信息通过滤波器进行特征提取后，放到深度学习模型中，然后让深度

基金项目：国家自然科学基金项目（62221003）

学习模型决定这一数据包（或者这一条流）应该遵循什么样的路由路径。在这一模型中，有的解决方案会选用卷积神经网络^[1]或循环神经网络^[2]，有的则会选择图神经网络^[3]。

在使用智能模块辅助路由计算的解决方案中，传统方法或启发式算法依然会被启用。例如，在一个典型的路由优化问题中，神经网络可能只是替代网络环境建模、拥塞检测或者流量预测中的一个模块，其他模块依然采用启发式算法来进行计算优化。例如：有的工作仅仅是利用神经网络对链路的拥塞情况进行预测，然后采用启发式算法决策^[1]；有的工作则对路径时延进行实时预测，然后基于预测的时延进行路由决策^[2]。

1.2 基于强化学习的技术

强化学习是一种并不需要标记网络状态及其对应策略，而是提供一个路由环境供强化学习智能体进行学习的技术。在每个时间点 t ，强化学习的智能体都能观察到当前的状态 s_t ，并相应做出行动 a_t ，且收到反馈奖励 r_t 的过程。强化学习智能体的任务是，不断地尝试最大学到的奖励 r_t 的累积求和。在采用强化学习进行路由优化的工作中，通过采取的算法，我们也可以对现有工作进行简单区分。现有强化学习算法要么采取简单传统的Q-learning进行优化，要么采用深度神经网络来学习更加复杂的策略。这其中，Q-learning以及深度神经网络便是强化学习智能体的两种表示形式。

采用Q-learning优化的智能路由算法最早可以追溯到1994年。BOYAN J. A.等提出将路由转发过程建模为马尔可夫决策的过程^[4]，以路由下一跳即将选择的节点作为动作，路由每一跳所花费的时延作为奖励值。通过对每一条延迟进行优化的方式，可以有效避免路由造成的网络拥塞的发生。

采用深度神经网络的强化学习方法在输入、输出以及奖励函数上可能和上述的Q-learning对应的表示方法一致。其主要差别在于，上述Q-learning可能基于表等简单的数据结构对状态-动作的映射进行记录，但深度神经网络会采用复杂的神经网络对这一映射进行表示。XU Z.等的工作便采用多层的神经网络对智能体的策略进行表示^[5]。

2 基于机器学习的智能路由主要场景

除了上述按照算法来分类外，我们还可以按照场景来对智能路由进行分类。近年来，在数据中心网络、无线网络中，智能路由的解决方案都广泛地被研究者们提出。这主要是因为，数据中心网络和无线网络有一个突出的特性，即网络内的设备大多由同一组织、实体等控制，更新换代比较容易，同时也便于部署集中式算法。对于广域网传输等领域，

由于涉及的参与者众多（无线路由器、骨干网路由器、接入网路由器很可能各自归属不同的厂商），很难在短期内迅速部署一些新的路由算法。因此，本节中，我们主要围绕数据中心网络和无线网络，特别是无线传感器网络，介绍智能路由近年来的一些代表性进展。

2.1 数据中心中的智能路由

数据中心是智能路由较容易应用的一类场景。在数据中心中，发送端、接收端、网内路由器交换机等均由同一实体控制，便于运营者集中化、中心式管理。同时，由于网络规模小、设备易于控制，设备可以轻易更新为面向集中控制的路由器，例如支持软件定义网络的路由器等。在部署上，各方面都为智能路由的部署创造了条件。

例如，文献[6]提出通过深度强化学习来对全局路由表进行优化，然后再通过软件定义网络将每个路由器的路由表部署到相应路由器上，可以带来智能路由性能提升。文献[7]进一步将智能路由技术与多路径等技术进行结合，提出了基于强化学习的数据中心多路径路由技术。这些技术都在数据中心场景下，采用强化学习等智能路由技术对路由算法进行了优化，并在实际部署中取得了一定的效果。

2.2 无线网络中的智能路由

无线网络一方面作为接入网，为用户提供接入互联网的服务，似乎不需要路由；另一方面，在传感器网络等场景中，也有许多路由操作需要完成。相比于有线网络，无线网络更为复杂的是：没有固定的有线链路——理论上只要在覆盖范围内的两个节点都可以建立链接通信。这为路由算法的优化引入了新的复杂度：路由算法需要与网络拓扑进行协同优化。基于深度学习等智能路由技术恰好能够解决这一复杂的问题，其在无线场景的应用也因此应运而生。

无线网络路由问题的优化目标可能更为复杂：除去基本的延迟、吞吐等性能优化外，很多场景的应用都需要注重路由算法的能耗、健壮性等。目前已经有了一些初步的尝试，如文献[8]就介绍了一种可以让无线传感器网络中的路由结果变得更加节能、健壮的机器学习算法。基于不同应用场景，运营者可以很轻易地对优化目标做修改，重新训练机器学习模型便可获得一个目标不同的智能路由模型，因此这一场景下的智能路由模型具有很强的迁移性。TANG F.等^[9]提出，基于深度学习的算法可以整合路由与拓扑优化，直接进行全局优化。通过构建实时的深度学习流量控制机制，该工作可以直接对各个节点发送流量以及目的节点进行优化控制，达到整体上降低节点间发送延迟、提高吞吐量的目的。

3 基于超图的智能路由解释技术

本节中，我们介绍一种使用超图来对基于机器学习的智能路由算法进行解释的技术。我们首先简要介绍超图，以及如何使用超图来表示网络系统。然后，介绍如何使用超图来解释网络系统中的关键组件。

3.1 超图形式化

超图由点和超边组成。一个普通图中的边与超图中超边的主要区别在于：超边可以覆盖多个顶点，如图1所示。我们将所有顶点和超边的集合分别表示为 V 和 E ，每个顶点 v 和超边 e 的特征分别表示为 f_v 和 f_e ，所有顶点和超边的特征矩阵分别表示为 F_V 和 F_E 。

使用超图，我们还可以表示SDN路由优化的建模。SDN控制器收集所有数据平面交换机的信息。在这种情况下，SDN路由优化器分析每个源-目的地对的流量需求，并基于拓扑结构和链路容量为所有源-目的地流量需求生成路由路径。

然而，由于路径中交换机的数量是不确定的，因此路由路径是很难表示的高阶信息。以前研究者尝试使用整数规划来表示路径，但是这种方法很难在有限的时间内进行有效优化。RouteNet^[3]设计了一种基于深度神经网络的优化算法，以连续地为每个源-目的地流量需求对选择最佳路由路径。

为了使用超图来形式化系统，我们将路径视为超边，物理链路视为顶点。覆盖顶点的超边表示该对需求的路径包含一条链路。图1展示了超图映射结果的一个例子。链路 $(1, 2, \dots, 8)$ 被建模为顶点。两对传输需求 $(a \rightarrow e)$ 和 $a \rightarrow g$ 被建模为超边（表示为 e_1 和 e_2 ）。顶点特征 F_V 是链路容量。超边特征 F_E 是每对交换机之间的流量需求量。如果超边 e 覆盖顶点 v ，则 e 的对应的网络流量需求应通过 v 的对应的链路。

RouteNet生成整体路由结果，即所有流量需求的路径。例如，假设RouteNet决定从 a 到 e 的需求通过链路 $2, 5, 6, 8$ （蓝色路径），从 a 到 g 的需求通过链路 $1, 3, 6, 8$ ，相应的超图应该是图1（c）。超边 e_1 覆盖顶

点 $2, 5, 6$ ，超边 e_2 覆盖 $1, 3, 6, 8$ 。

所有顶点-超边连接 $\{v, e\}$ 是：

$$\{(2, e_1), (5, e_1), (6, e_1), (1, e_2), (3, e_2), (6, e_2), (8, e_2)\}, \quad (1)$$

其中，解释工作的关键是我们需要知道哪些是对整体路由决策至关重要的

连接。如果网络管理员知道这个机器学习的智能体中决定性的决策，那么他们或许就可以从关键决策的对比中理解智能路由决策的原因。在这里，我们总结了我们能用超图对路由结果表示的两个原因：

1) 图结构化输入或输出。由于图是超图的一种简单形式，如果全局系统的输入或输出是图结构化的，则该系统可以自然地用超图表示。

2) 二元变量映射。如果全局系统构建两个变量之间的映射，这两个变量可以用顶点和超边表示。映射可以用超图中的连接关系表示。这不仅包括路由，有时候也可以拓展至许多资源分配系统中资源（如物理服务器）和请求（如网络功能）之间的映射构建。

只要全局系统具有上述特征之一，我们就可以用超图进行表示并解释。

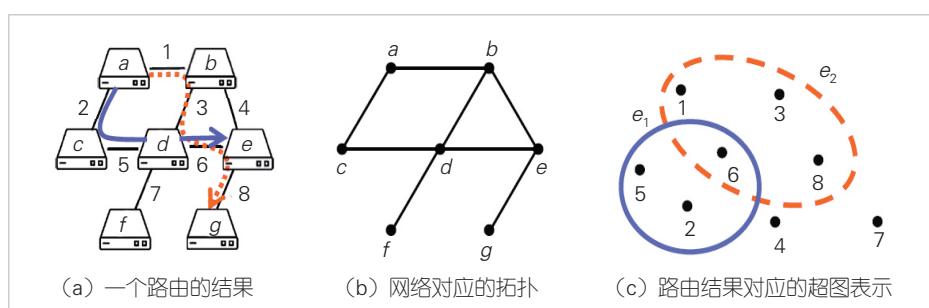
3.2 关键连接搜索

下一步，我们要找到对原始系统优化结果至关重要的顶点-超边连接。首先，我们介绍超图的关联矩阵表示。关联矩阵 I （大小为 $|E| \times |V|$ ）是一个0-1矩阵，用于表示顶点和超边之间的连接关系。 $I_{ev} = 1$ 表示超边 e 包含顶点 v 。例如，图1中超图的关联矩阵为：

$$I = \begin{pmatrix} 0 & 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 & 1 \end{pmatrix}. \quad (2)$$

我们的设计目标是评估原始系统的每个连接对优化结果的影响。以SDN路由为例，我们将评估式（2）中每个（链接、路径）连接对原始系统优化结果的影响。我们通过分数的关联矩阵 $W \in [0, 1]^{|\mathcal{E}| \times |\mathcal{V}|}$ 来表示每个超边-顶点连接的重要性。如果 v 和 e 之间没有连接，则 $W_{ev} = 0$ 。关键连接搜索算法的概述为：

$$\min D(Y_w, Y_l) + \lambda_1 ||W|| + \lambda_2 H(W) \quad \text{s.t. } 0 \leq W_{ev} \leq I_{ev}, \quad \forall v \in V, e \in E, \quad (3)$$



▲图1 用超图表示的路由结果

其中, $D(Y_w, Y_l)$ 、 $\|W\|$ 、 $H(W)$ 分别展开为:

$$D(Y_w, Y_l) = \begin{cases} \sum Y_w \log \frac{Y_w}{Y_l}, & \text{离散} \\ \sum \|Y_w - Y_l\|^2, & \text{连续} \end{cases} \quad (4)$$

$$\|W\| = \sum_{v,e} |W_{ev}|, \quad (5)$$

$$H(W) = -\sum_{v,e} (W_{ev} \log W_{ev} + (1 - W_{ev}) \log (1 - W_{ev})), \quad (6)$$

其中, 优化目标包含3个部分:

1) 性能退化 $D(Y_w, Y_l)$ 。关键的连接应是那些对网络系统输出有很大影响的连接, 这是与任务无关的。因此, 我们需要测量机器学习智能路由系统原始输入特征和掩码 W 加权的输入特征的差别。掩码特征(需求、容量)生成的路由决策应该与原始决策类似。我们将原始输入和输入掩码 W 生成的决策分别记为 Y_l 和 Y_w 。因此, 我们最大化 Y_w 和 Y_l 之间的相似性, 记为 $D(Y_w, Y_l)$ 。路由系统的输入/输出有时是连续的, 有时也可能会是离散的。例如, 当预测一条流的转发路径时, 路由系统的输出就是一个离散的节点序列; 当预测一条链路上的拥塞程度时, 路由系统的输出就是一个连续的值。针对不同的变量类型, 我们分别采用不同的距离估计方法。我们采用KL散度来测量离散输出(如路由决策序列)和均方误差。这两者都是机器学习社区中常见的相似性度量。

2) 解释简洁性 $\|W\|$ 。通常, 人们可以理解的解释数量是有限的。因此, 关键连接的数量也应尽可能少, 以便网络操作员理解。如果算法提供了太多的“关键”连接, 网络操作员则会感到困惑, 无法轻松解释网络系统。我们将 W 的简洁性定义为所有元素的和(矩阵的规模)。另外, 我们还需要在优化目标中惩罚掩码 W 的大小。

3) 确定性 $H(W)$ 。此外, 我们还期望 W 的结果是确定的, 即对于每个连接 (v, e) , 要么和结果毫无关系(W_{ev} 接近 0), 要么对结果影响很大(W_{ev} 接近 1)。否则, 智能体可能将学会掩盖所有具有相同权重的连接并生成无意义的解释。在本文优化掩码 W 的熵, 以鼓励 W 中的连接接近 1 或 0, 其中熵是信息论中的不确定性度量。

为了平衡上述的优化目标,

我们为网络操作员提供了两个可定制的超参数(λ_1 和 λ_2)。例如, 路由结果的在线监视器可能只需要最关键的信息来帮助它快速做决策, 而如果管理员想要离线分析解释结果, 则解释结果必须要很详细才能进一步改进智能路由模型。在这种情况下, 网络操作员可以增加(或减少) λ_2 来减少(或增加) 具有中位掩码值的未确定连接的数量。然后, 我们将向网络操作员公开更少(或更多)的关键连接。网络操作员可以在其应用程序中调整优化目标不同部分的权重。

在这种情况下, 连接对输出的贡献可以被定量表示。我们可以通过连接所涉及的链路流量来定量地了解连接对输出的贡献。同时, 还可以通过掩码值来进一步判断哪个连接上的哪个流量在整体结果中起主导作用。这样, 我们就可以提供更细粒度的解释。

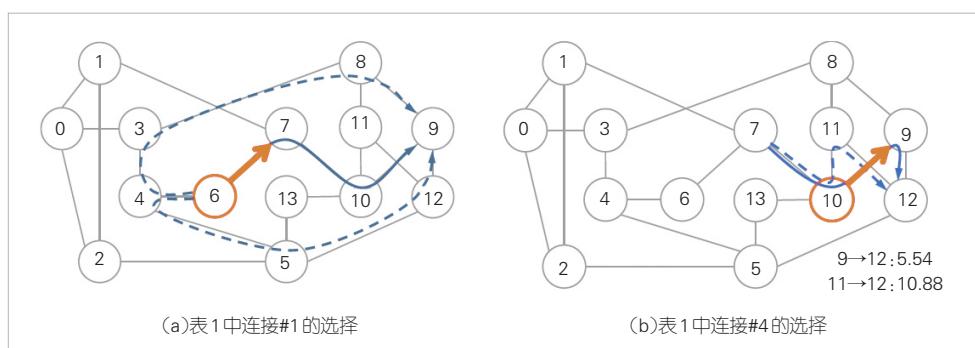
4 一个解释的例子

我们首先复现了 RouteNet^[3]的代码原型, 并基于真实流量数据在 RouteNet 上测试了所提出的解释方法。同时, 我们测试了 RouteNet 中的 NSFNet 拓扑以及流量数据, 具体如图 2 所示。如第 3 节中所描述的解释方法, 每个连接的掩码值 W 代表了该连接对于最终智能路由方法的决策影响程度。掩码值前 5 名的连接组合列见表 1。

在表 1 中, 可以看出, 掩码值越接近于 1, 该链路的选择对整体路由路径的选择就越关键。例如, 路由路径 6→7→10→9 和链路 6→7 之间的掩码值为 0.886, 这说明在节点 6

▼表1 RouteNet 在 NSFNet 上的解释结果的前 5 名掩码值

| | 路由路径 | 关键链路 | 掩码值 M_{ve} | 决策原因 |
|----|-----------|------|--------------|------|
| #1 | 6→7→10→9 | 6→7 | 0.886 | 路径更短 |
| #2 | 1→7→10→9 | 1→7 | 0.880 | 路径更短 |
| #3 | 7→10→9→12 | 10→9 | 0.878 | 负载更低 |
| #4 | 8→3→0→2 | 8→3 | 0.875 | 路径更短 |
| #5 | 6→4→3→0 | 6→4 | 0.874 | 负载更低 |



▲图2 两个RouteNet路由结果解释方法的示意

上做出选择链路 $6\rightarrow 7$ 这一个决策十分关键。与之相反，在节点7上选择链路 $7\rightarrow 10$ 、在节点10上选择链路 $10\rightarrow 9$ ，可能因为这些结果的选择都比较显然，或者它们对性能影响不大，因此就不是那么关键。例如，在节点7上如果不选择 $7\rightarrow 10$ 而选择 $7\rightarrow 1$ ，那么便和目的节点南辕北辙了，选择 $7\rightarrow 1$ 的路径性能会显著劣于 $7\rightarrow 10$ ，因此做出这一判断并非难事。

我们选取了表1中路径更短、负载更低这两个原因中掩码值最高的两条路由路径分别做进一步的解释。

对于#1，如图2（a）所示，如果流量想从节点6（源）到节点9（目的），路由算法选取的路径就有很多，常见的有3条（如图2（a）中蓝色的3条）。这3条路径长度差不多，均不超过4跳；而其他路径则都要在5跳以上才能到达节点9。在3条候选路径（蓝色）中，最短路径（实线路径）的第一跳是 $6\rightarrow 7$ ，而其他路径（虚线路径）的第一跳是 $6\rightarrow 4$ 。因此，我们发现，第一跳的选择对于从6到9的路径来说是重要的，因此选择 $6\rightarrow 7$ 。一旦路由算法能够正确地将第一跳的路由决策选择为 $6\rightarrow 7$ 而不是 $6\rightarrow 4$ ，那么后面的路由决策就变得很简单了。因此，第一跳的决策在路由路径中至关重要。

对于#2，如图2（b）所示，如果流量想要从节点7（源）到节点12（目的），长度相等的最短路径有2条（图2（b）中蓝色的两条），网络管理员需要理解为什么RouteNet选择了 $7\rightarrow 10\rightarrow 9\rightarrow 2$ 。我们的解释方法显示，在节点10上时，选择 $10\rightarrow 9$ 这个链路而非 $10\rightarrow 11$ 非常重要。有了这个线索，我们在进一步分析两条路径的负载情况后发现， $11\rightarrow 12$ 链路的负载大约是 $9\rightarrow 12$ 的2倍（如图2（b）所示）。如果要避开 $11\rightarrow 12$ 的链路，到了节点11再避开就太晚了，因为次优的路径将不得不选择 $7\rightarrow 10\rightarrow 11\rightarrow 8\rightarrow 9\rightarrow 12$ ，远远长于 $7\rightarrow 10\rightarrow 9\rightarrow 2$ 。因此，在节点10上选择链路 $10\rightarrow 9$ 这一决策对性能影响非常大。我们的解释算法能够准确揭示出来这两个决定，显示了解释方法的有效性。

5 结束语

随着基于机器学习的智能路由技术的广泛应用，对这些机器学习技术进行解释的需求也会应运而生。这可以让更多的智能路由技术得到应用，推动智能路由的部署。随着机器学习技术的发展以及可解释性的提高，性能更强的智能路由方法将会出现，路由算法的提升前景将会非常乐观。未来，

智能路由技术及其解释技术将会成为网络路由的主流技术之一。

参考文献

- [1] BARABAS M, BOANEA G, DOBROTA V. Multipath routing management using neural networks-based traffic prediction [EB/OL]. [2023-08-12]. http://personales.upv.es/thinkmind/EMERGING/EMERGING_2011/emerging_2011_6_30_40129.html
- [2] VALADARSKY A, SCHAPIRA M, SHAHAF D, et al. Learning to route [C]// Proceedings of the 16th ACM Workshop on Hot Topics in Networks. ACM, 2017: 185–191. DOI: 10.1145/3152434.3152441
- [3] RUSEK K, SUAREZ V J, MESTRES A, et al. Unveiling the potential of Graph Neural Networks for network modeling and optimization in SDN [C]//In proceedings of the ACM symposium on SDN research, ACM, 2019: 140–151
- [4] BOYAN J A, LITTMAN M L. Packet routing in dynamically changing networks: a reinforcement learning approach [C]//Proceedings of the 6th International Conference on Neural Information Processing Systems. ACM, 1993: 671–678. DOI: 10.5555/2987189.2987274
- [5] XU Z Y, TANG J, MENG J S, et al. Experience–driven networking: a deep reinforcement learning based approach [C]//Proceedings of IEEE INFOCOM 2018 – IEEE Conference on Computer Communications. IEEE, 2018: 1871–1879. DOI: 10.1109/INFOCOM.2018.8485853
- [6] LIU W X, CAN J, CHEN Q C, et al. DRL-R: deep reinforcement learning approach for intelligent routing in software-defined data-center networks [EB/OL]. [2023-08-10]. <https://www.sciencedirect.com/science/article/pii/S1084804520303313>
- [7] JUA A, SINGH K K , VIMALA D K, et al. Reinforcement learning based weighted multipath routing for datacenter networks [EB/OL]. [2023-08-10]. <https://www.sciencedirect.com/science/article/pii/S2214785321003412>
- [8] NAYAK P, SWETHA G K, GUPTA S, et al. Routing in wireless sensor networks using machine learning techniques: challenges and opportunities [J]. Measurement, 2021, 178: 108974. DOI: 10.1016/j.measurement.2021.108974
- [9] TANG F X, MAO B M, FADLULLAH Z M, et al. On removing routing protocol from future wireless networks: a real-time deep learning approach for intelligent traffic control [J]. IEEE wireless communications, 2018, 25(1): 154–160. DOI: 10.1109/MWC.2017.1700244

作者简介



孟子立，清华大学在读博士研究生；主要研究领域为实时多媒体传输与基于人工智能的网络系统；曾获微软学者奖学金、字节跳动奖学金等；发表论文30余篇。



徐明伟，清华大学教授、网络科学与网络空间研究院执行院长、国家杰出青年基金获得者、国家自然科学基金委创新群体负责人、国家“万人”计划领军人才、国家重点研发计划“宽带通信与新型网络”和“多模态网络与通信”重点专项专家、中国通信学会常务理事、中国计算机学会互联网专委会副主任委员；主要研究领域为互联网体系结构、大规模路由和网络空间安全。



网络智能传输研究进展

Network Intelligent Transmission Technology: A Survey

廖乙鑫/LIAO Yixin¹, 王子逸/WANG Ziyi¹, 崔勇/CUI Yong²

(1. 北京邮电大学, 中国北京 100876;

2. 清华大学 中国 北京 100084)

(1. Beijing University of Posts and Telecommunications, Beijing 100876, China;

2. Tsinghua University, Beijing 100084, China)

DOI: 10.12142/ZTETJ.202305010

网络出版地址: <http://kns.cnki.net/kcms/detail/34.1228.TN.20231016.0846.002.html>

网络出版日期: 2023-10-17

收稿日期: 2023-08-10

摘要: 新兴超低时延场景的出现以及6G技术与人工智能技术的发展,促使网络智能传输成为研究热点。分析了传输层和应用层的时延组成及影响因素,对机器学习技术与传输层、应用层流媒体传输相结合的智能传输协议的发展和优缺点进行了综述。从传统网络传输协议的发展、人工智能技术的发展、网络传输和人工智能结合3个方面展望了网络智能传输面临的机遇与挑战。认为分布式机器学习训练场景的传输性能、训练数据的质量、模型的泛化能力、模型大规模部署的开销是未来网络智能传输技术的重点研究方向。

关键词: 网络传输; 机器学习; 时延分析; 拥塞控制; 视频传输

Abstract: With the emergence of emerging ultra-low latency scenarios and the development of 6G technology and artificial intelligence technology, intelligent network transmission has become a research hotspot. The delay components and influencing factors of the transport layer and application layer are discussed. Then, the development, advantages, and disadvantages of intelligent transmission protocols that combine machine learning technology with transport layer and application layer streaming media transmission are reviewed. The opportunities and challenges faced by intelligent network transmission are prospected from three aspects: the development of traditional network transmission protocols, the development of artificial intelligence technology, and the combination of network transmission and artificial intelligence. It is believed that the transmission performance of distributed machine learning training, the quality of training data, the generalization ability of the model, and the cost of large-scale deployment of the model is the key research direction of future network intelligent transmission technology.

Keywords: network transmission; machine learning; analysis of delay; congestion control; video transmission

引用格式: 廖乙鑫, 王子逸, 崔勇. 网络智能传输研究进展 [J]. 中兴通讯技术, 2023, 29(5): 61–67. DOI: 10.12142/ZTETJ.202305010

Citation: LIAO Y X, WANG Z Y, CUI Y. Network intelligent transmission technology: a survey [J]. ZTE technology journal, 2023, 29(5): 61–67. DOI: 10.12142/ZTETJ.202305010

随着6G技术与人工智能技术的迅速发展,我们正面临着众多机遇和挑战。6G技术提供了低时延的物理传输通道^[1],为扩展现实(XR)、远程医疗、元宇宙、车联网等新兴应用场景^[2-4]开拓了广阔的发展空间。然而,新兴应用场景的发展还需要上层传输层、应用层的架构完善来支撑。新兴应用与传统应用在时延要求^[5]、承载网络结构等方面存在明显差异,这对传统网络传输协议提出新的挑战。传统的传输层协议和应用层协议设计通常基于静态的规则和网络假设,难以适应现代网络中的复杂变化和多样化应用需求。借助人工智能技术^[6],传输协议可以通过学习来适应实时网络

环境,调整拥塞控制和传输优化策略。比如,机器学习^[7]算法能够从大量的复杂网络数据中学习并提取有用的模式和特征,动态生成网络传输协议规则。机器学习技术还具备适应的能力,能根据网络环境的变化调整模型参数和优化算法。将人工智能技术应用于传输协议的设计,可以更好地应对网络中的动态变化和多样化应用需求,提供更高效、可靠和稳定的网络传输。网络智能传输技术应运而生。

在网络传输中,时延是一个非常重要的指标,对新兴应用场景而言尤为关键。为了更准确地为网络智能传输技术设定学习的优化目标^[8],提供高质量的传输体验,我们有必要分析传输层协议和应用层协议的时延形成机制和影响因素^[9]。对于传输层而言,主要的底层协议为传输控制协议(TCP)和用户数据报协议(UDP)。TCP提供面向连接的传

基金项目: 自然科学基金重点项目(62132009); 创新研究群体项目(62221003)

输服务，因而会引入连接管理、可靠数据传输、拥塞控制相关的时延；UDP提供尽力而为的传输服务，因而只引入数据传输的时延。对于应用层而言，时延的组成和影响因素由应用类型决定。我们以车联网^[10]和低时延直播场景^[11]为例，分析两种场景的时延组成和影响因素。为了设计更符合场景需求的传输协议，减少传输时延，需要根据具体协议的时延影响因素，设置准确恰当的智能优化目标。

学界对机器学习算法应用于传输层协议和网络层协议进行了长期探索，并取得了显著成果。对于传输层，拥塞控制算法构成了传输层协议的核心组成部分，其优秀性能体现在合理调节数据端点的发送能力。利用机器学习技术理解复杂网络的流量等特性，并针对性地设计拥塞控制算法，提升其鲁棒性、泛化性、训练效率和推理效率，是传输层研究的重点。对于应用层，由于视频流量占据互联网的大部分流量，如何合理地利用机器学习技术设计高效率的视频码率自适应（ABR）算法^[12]，提升视频质量和用户视频体验质量（QoE）^[13]，并能恰当地在现实中进行部署，是学界的重要研究方向。

网络智能传输技术的研究取得了显著成果，但依然面临一些挑战。快速 UDP 网络连接（QUIC）^[14-16]、截止日期感知传输协议（DTP）^[17]等新兴协议的提出，Media Over QUIC^{[18][19]}工作组的成立，进一步推动了网络传输协议研究的发展。同时，AI 大模型的发展，为设计和训练更精确的网络大模型带来新的可能性。然而，很多问题的解决方案依旧亟待探索，例如：构建与现实世界分布一致的网络训练数据集，充分利用其中的带宽、时延等指标训练优秀的网络传输模型；设计网络传输协议，解决分布式机器学习的节点高效通信问题，在部署机器学习模型时，适应不同上层应用对网络传输指标的不同需求，并在网络动态变化时保持良好的泛化性。

本文对广域网点对点网络智能传输研究进展进行了综述：首先对传输层、应用层的时延成因进行详细分析；其次介绍了传输层、应用层的智能传输研究现状，包括人工智能与拥塞控制算法相结合、人工智能与流媒体传输相结合的成果进展；最后对网络传输新发展、人工智能技术的新发展、网络传输和人工智能结合这 3 个方面进行展望，并预测了未来的研究发展趋势和可能存在的挑战。

1 时延的形成机制

为了提高网络协议的传输质量，最大程度减少时延，有必要先讨论时延的种类和影响因素。网络层的时延主要包括分组传输时延和分组处理时延，受网络拥塞程度、路由选

择、路由器缓存和处理能力影响^[20]。数据链路层的时延主要包括帧传输时延和帧处理时延，受链路带宽、帧长度以及链路错误率等影响^[21]。本文重点介绍传输层和应用层的时延影响因素。

1.1 UDP 时延分析

UDP 是无连接协议，它的时延由传输时延、传播时延、排队时延、处理时延组成。由于 UDP 本身的无连接性和不可靠性，它不需要处理 TCP 中的握手时延、挥手时延、重传时延，也不进行拥塞控制^[22]。总的来说，UDP 的时延通常少于 TCP。

1.2 TCP 时延分析

在 UDP 协议的基础上，TCP 能够确保可靠的端到端传输。然而，TCP 的使用会引入多种类型的时延。TCP 连接的生命周期包括三次握手、数据传输和四次挥手过程。这些过程会涉及连接的建立和断开、数据的传输和传播、数据在中间端点的排队和处理，以及协议本身可能引发的队头阻塞、重传、拥塞控制等。以上过程均涉及对端通信过程中的时延开销。

传输时延是指数据包发送到链路中的时延，取决于数据包的大小以及网络的带宽。传播时延依赖于两端的物理距离和传输介质。排队时延表示数据包在网络设备的缓冲区中排队时的时延。在数据包到达目标地后，网络设备需要将数据包从接收缓冲区中取出并进行处理，这将引入处理时延。对于新建立的 TCP 连接，拥塞控制算法会控制数据包注入网络的速度^[23]，因而会引入拥塞控制时延。这会受到拥塞控制算法的预设规则的影响（如慢启动参数）。此外，TCP 保证数据有序交付的特性会引入队头阻塞时延，该时延可能由数据包分组丢失或数据包无序到达引起。当分组丢失或超时到达时，重传的数据包同样也会引入重传时延^[24]。

1.3 应用层时延分析

应用层的时延影响因素由应用类型决定。以车联网场景为例，对于道路自动驾驶而言，车载计算硬件需要在毫秒级别内识别道路中的行人和障碍。在这个过程中，存在传感器数据的传输时延以及计算硬件的计算和决策时延。汽车需要实时与红绿灯等交通设施通信，并与云交换驾驶信息，以优化全局交通流量，这涉及云的决策和处理时延^[25]。在交互式视频直播场景里，主播与观众之间的端到端时延可以划分为 4 个部分，分别是主播端视频内容上传到服务器的时延、将视频从服务器下载到客户端的时延、客户端的播放缓冲时延

和这个过程中涉及的编解码时延^[26]。

2 传输层研究现状

拥塞控制算法是传输层协议的核心。良好的拥塞控制算法应该合理地控制端点发送数据的能力。拥塞控制方案的性能会受到许多因素的影响，包括流量模式、链路故障、动态延迟、数据包丢失等。针对复杂网络环境很难设计最优甚至接近最优的预定义静态规则的策略。通过人工智能技术，计算机可以生成更灵活、表现更优良的拥塞控制策略。拥塞控制算法的研究时至今日一直在进行中。本小节将介绍设计传输层协议的网络智能传输研究成果。

随着机器学习技术的发展，研究人员发现不同的机器学习训练策略效果有显著差别。2013年，WINSTEIN K.等提出Remy^[27]，旨在生成用于各个通信端点的拥塞控制算法。Remy通过离线学习拥塞控制规则来执行优化，将预设定的网络条件、优化目标（如高吞吐量和低排队时延）作为目标函数输入，并通过离线学习来优化目标函数，最终生成一个规则表来指导拥塞控制操作。Remy离线优化后不再进行进一步学习，因此无法泛化到不同的网络场景。当前流量/网络条件偏离Remy的网络输入假设可能会导致性能不佳。

SIVARAMAN A.^[28]等对基于数据驱动的端到端拥塞控制算法的设计过程，进行了形式化讨论，量化了训练符合要求的模型的难易程度。此后，JAY N.采用了基于线性奖励函数的深度强化学习方法，提出了拥塞控制框架Aurora^[29]，并证明了其训练深层神经网络以捕获复杂流量和网络条件模式的能力。该框架在拥塞控制性能上能够与最先进技术相媲美，甚至优于它们。2022年，HUANG T. C.针对互联网短视频上传的场景，提出了一种名为DuGu的拥塞控制框架^[30]。该算法通过模仿最优求解器而不是使用手工设计的奖励函数，从而提高性能和学习效率。同时，这种方法降低了过度依赖奖励函数导致的学习策略不稳定行为的影响。

直接将AI模型与网络传输结合，可能会遇到推理效率、训练效率的问题。DONG M.提出面向性能的拥塞控制框架(PCC Vivace)^[31]。Vivace将机器学习的凸优化理论应用于模型训练中，并将RTT梯度加入目标函数，来保证多个Vivace流可以收敛到一个公平和有效率的状态中，在随机丢包与收敛速度之间实现了有效均衡。此外，DONG M.使用了基于梯度的no-regret在线优化算法来调整发送速率，证明了在正确选出包含吞吐量、损耗和延迟的效用函数时，稳定的全局速率配置(纳什均衡)始终存在。ZHANG L.提出截止时间感知传输机制D3T，利用深度强化学习算法来决定发送速率和前向纠错(FEC)冗余率^[32]，旨在提高用户体验质量

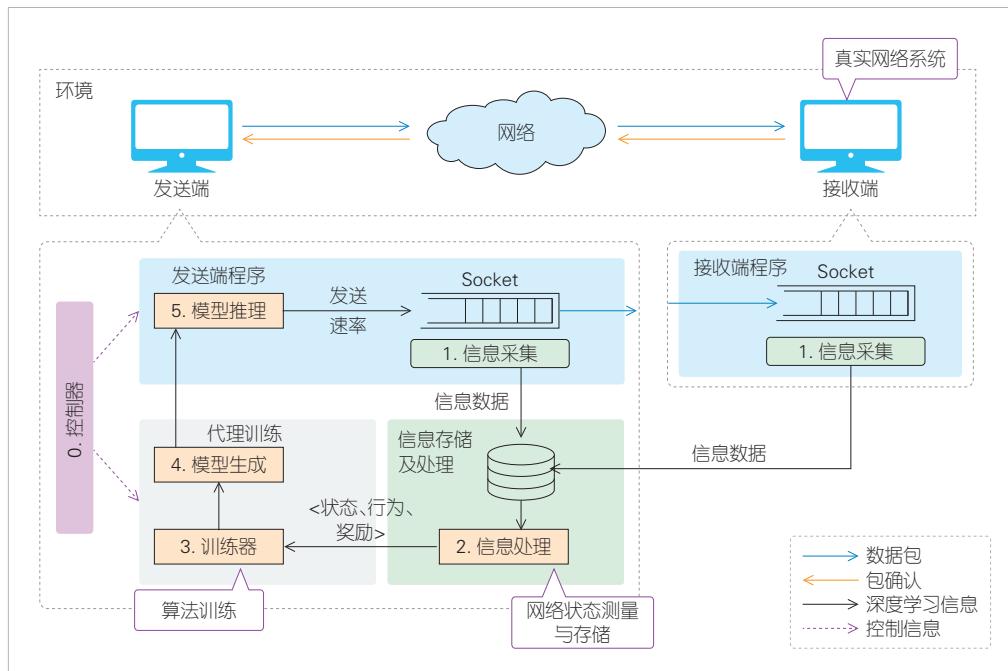
(QoE)以满足截止时间要求。为了减小训练强化学习决策代理的难度，该研究设计了一种启发式调度算法，用于训练FEC和拥塞控制模块的强化学习代理。

由于网络流量的特征差异，训练后的机器学习策略在处理异常流量时，可能会面临严重的鲁棒性、泛化性问题。2019年，NIE X. H.^[33]提出TCP-RL，根据在Web服务器端观察到的实时网络状况，为短流动态配置适合的初始窗口大小(IW)，并通过演员评论家算法(A3C，一种深度学习算法)为长流动态配置合适的拥塞控制(CC)策略。将深度学习技术与传输协议的静态配置结合，可以增强协议的稳定性、鲁棒性，进一步降低智能传输协议的训练、推理成本。SOHEIL A.在2020年提出了一种名为Orcal^[34]的混合拥塞控制机制，通过结合深度强化学习技术和传统拥塞控制机制，解决了应用深度学习技术可能遇到的高计算开销、低泛化性和过度收敛等问题。Orcal在洲际、洲内和蜂窝环境等复杂的网络场景中经过性能测试，并取得了良好的性能表现。因此，引入两级控制机制可以显著缓解未见场景中的性能问题。

模拟或仿真平台无法完全仿现现实世界网络流量的多变性等特性。ZHANG L.提出了拥塞控制策略学习框架ARC^[35]，如图1所示。该框架在真实环境中进行异步执行，通过在发送端和接收端分别采集训练所需的网络状态，并在信息存储与处理模块中构建强化学习训练所需的数据元组，解决了训练数据的异步问题。ARC框架通过设计模型训练与真实网络系统异步运行、模型推理与数据发送异步运行，解决了模型的动态更新问题，并取得了高吞吐量和低延迟的良好性能。随后，ZHANG L.提出了一种名为Deep CC的学习训练架构^[36]。Deep CC利用多目标深度强化学习算法训练深度神经网络模型。考虑到动态网络条件下应用的特定需求，Deep CC增加了在线微调功能，以提升其在应对应用程序需求变化和不同网络条件下的泛化能力。这使得在网络特征变化时，Deep CC不需重新训练就能够做出符合各种吞吐量、延迟和丢包率等均衡需求的决策。

3 应用层流媒体传输研究现状

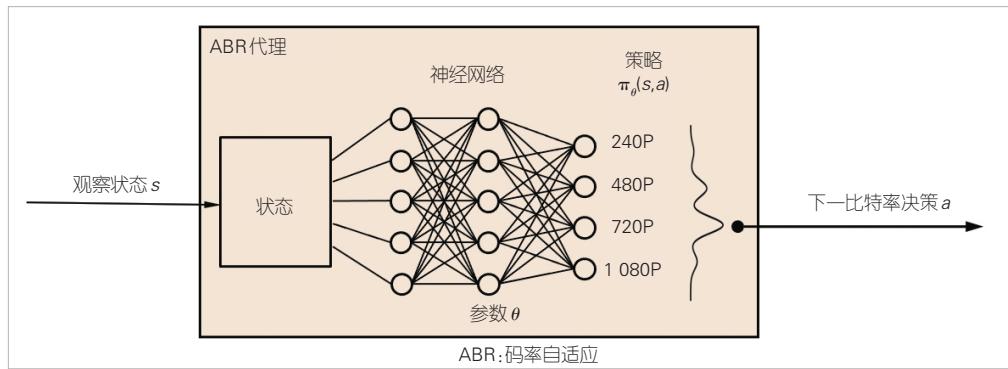
越来越多的学者开始研究将机器学习技术与传统应用层协议或算法相结合，以便更好地感知网络条件的复杂模式。近年来，视频流量快速增长，同时用户对视频质量的要求也在不断提高。ABR算法是内容提供商用来优化视频质量的主要工具，其准确性和性能会对视频流性能造成不可忽视的影响。然而，难以建模的网络变化和难以平衡的视频质量目标(如最大化比特率和最小化停顿)等因素，给设计高准确



▲图1 ARC学习框架结构

性、高效率的ABR算法带来困难。近年来，学界将人工智能技术应用于ABR算法，以增强其在底层网络变化时的适应性、效率和视频质量目标的平衡性能，并且取得了丰硕的成果。

2017年，MAO H. Z.提出Pensieve^[37]，使用A3C算法生成ABR算法，以提高广泛网络条件和不同QoE目标下的性能。Pensieve基于历史传输指标选择未来视频块比特率，从而不依赖于网络环境的假设。如图2所示，观察状态s包含做出比特率决策相关的信息，如吞吐量测量值。一个由多个卷积网络（CNN）组成的2层神经网络，将状态映射到下一个比特率适应策略的概率分布。Pensieve将目标设定为最大化总奖励，通过训练神经网络的参数 θ ，确定合适的比特率选择来提供更好的奖励。测试中，Pensieve的平均QoE提高了12%~15%。Pensieve能够深入学习特定网络场景的特征，



▲图2 Pensieve算法架构

并根据经验选择适当的比特率决策进行输出，从而优化策略。然而这是一种采用离线训练并在线部署的方式，网络状况的变化可能会对其性能造成影响。为了解决这个问题，Pensieve需要支持定期更新ABR算法以应对新数据的到达。不过这会增加一定的计算开销，并需要系统从少量数据中快速收敛到优秀策略。因此，如何平衡再训练频率和性能仍然是一个待讨论的问题。

Pensieve等系统的速率控制方法侧重提供高视频比特率。但随着编码比特率的提高，视频质量的提升呈现出

边际递减的效应。过分追求高比特率只会加大网络传输的压力，而视频质量提升甚微。HUANG T. C.提出了视频质量感知速率控制算法（QARC）^[38]，目标是在实时视频流场景的比特率控制中平衡视频质量、延迟和带宽资源。QARC的视频质量预测网络（VQPN）模块通过历史视频帧预测未来视频质量，以解决训练过程中的“状态爆炸”问题。其视频质量强化学习（VQRL）模块使用A3C训练神经网络，根据VQPN的输出和历史时间网络状态数据，来输出满足视频质量和低延迟需求的比特率。

相似比特率和视频质量未必会为每一个用户都带来最好的QoE体验。这是因为传统的或数据驱动的ABR算法在优化QoE模型时，通常假设所有用户都具有相同的偏好^[39-40]。LAI Z. Q.提出一种全景比特率适应技术。在虚拟现实（VR）

场景中，当带宽不足而出现比特率降低时，该技术可优先降低用户视野外的视野质量，进而减少对用户的QoE影响^[41]。ZUO Y. T.^[42]提出一种视频系统Ruyi。该系统由偏好感知的QoE模型和基于神经网络的ABR算法组成。偏好感知的QoE模型输出用户偏好，并将其作为ABR算法神经网络的输入之一，从而

得出最佳比特率预测，以最大化用户特定的 QoE，无须为不同用户重新训练模型。仿真实验表明，Ruyi 能够提升所有用户的 QoE，提升幅度高达 65.22%。

先前的 ABR 算法研究中，预测模块都将吞吐量或交付时间视为优化目标。然而，至今还没有相关工作定量地研究这两个目标的效果差别。此外，传统的 ABR 算法均只考虑了网络条件波动对吞吐量造成的影响，忽略了应用程序本身行为（如 On-Off 周期）^[43] 的影响。GERUI L. 等通过视频流测量平台来定量验证预测吞吐量和交付时间的所有影响因素的相关性，发现块大小和吞吐量之间存在强相关性，并深受播放器状态、播放器客户端所在平台信号强度、块索引的影响；通过分别比较多元线性回归、决策树与两种优化目标 4 种组合的预测误差，确定基于吞吐量的优化目标优于交付时间，并提出 Lumos^[43]。Lumos 基于过去 t 个块的最大吞吐量、交付时间、客户端连接类型，以及最后块的比特率、大小、索引等训练回归决策树，通过集成到现有 ABR 算法中，提高吞吐量的预测精度。

4 展望与挑战

网络智能传输技术至今已经取得长足的进展，但仍面临诸多技术挑战。本文中我们从网络传输新发展、人工智能技术新发展、网络传输和人工智能结合的角度进行介绍。

4.1 网络传输新发展

传输层协议也在不断发展，基于 UDP 的 QUIC 协议被提出，并成为正式的 RFC^[14-16]。SHI H. 提出了 DTP^[17]，并首次在数据块中提出截止时间、数据块优先级的概念；ZHANG J. 在 DTP 基础之上设计了调度程序和自适应冗余机制^[44]，以满足动态网络的多样化需求。

单路径的网络传输可能存在网络不稳定和吞吐量不足的问题。与单路径相比，多路径传输的优点是能够针对不同场景^[45] 和优化目标^[46] 提供网络的无缝切换和更大的聚合带宽。ZUO X. T. 提出了截止日期感知的多路径调度框架^[47]，根据网络状态信息对发送端缓冲区块进行排序，决定数据块的发送顺序和路径分配，以减少带宽资源的浪费。

此外，国际互联网工程任务组（IETF）成立了 Media Over Quic 工作组^[18-19]，致力于建立利用 QUIC 协议发布媒体协议的机制。

4.2 人工智能技术新发展

近年来，机器学习技术在不断发展，出现了深度学习、强化学习等分支，提供了更强大的解决复杂问题的能力。然

而，从网络通信的角度看，机器学习领域仍存在一些挑战。

如何构建分布比例合乎现实世界且标注正确的高质量数据集，同时提供一套方法进行验证，是训练模型场景中的挑战。例如，将机器学习应用到传输层拥塞控制策略时，由于网络流量的多变性，如何确保采集到的网络流量数据符合现实网络场景的流量分布，依旧是构建网络智能传输模型的难点。

近年来，自监督表示学习在自然语言处理领域获得了惊人的成功，成为最近一些大语言模型的基础（如 LLaMA、LaMDA、Bard）^[48-49]。大量参数和海量训练数据是大模型取得惊人成功的原因。然而，如何合理地对海量的网络数据（如抖动、带宽、时延）进行标注再训练，并应用于网络传输中，是一个主要挑战。

4.3 网络传输和人工智能结合

将机器学习技术应用于网络传输是人们持续关注的研究领域。通过机器学习技术对特定网络的深层规则进行建模，可能会在决策效果上超越传统协议。然而，如何更好地将网络传输与 AI 模型结合仍然面临一系列的挑战。

分布式机器学习通过将训练任务拆解并部署到多个节点，来加快模型的训练速度。在实际部署中，节点之间的通信方式、通信性能会对整体训练性能造成很大差异，传输性能优化是热门的研究方向之一^[50-52]。如何根据分布式机器学习训练场景的数据传输周期性、数据中心流量对拥塞的敏感性的特点，及时避免拥塞，缓解交换机缓冲区的积压问题，依旧需要进一步研究。

对上层应用而言，不同的应用对网络传输的质量需求有差异，例如：实时交互性应用更关注抖动和延迟，而点播类应用更关注带宽。此外，对流媒体传输场景而言，不同用户对 QoS 的评判标准也有不同^[42]。为每一个用户和应用都训练一个决策模型是不现实的；反之，一个统一的决策模型难以在多样的传输指标中实现平衡。对底层网络而言，由于实际网络环境是动态变化的，基于离线训练和线上部署的模型需要具有足够的泛化性。如何设计出适当的目标函数、训练框架，在底层网络的延迟、丢包率等指标变化时，做出准确的决策，并且服务好上层应用不同的需求，将对网络传输模型的设计提出更高要求。

网络智能传输模型的决策具有实时性的特点，例如需要根据动态变化的带宽、时延等网络数据做出决策。在现实网络中部署时，传统的决策结果从服务器的神经网络模型决策获得，集群计算资源的不足将会影响模型推理的时间。为了能在网络中大规模部署，模型的推理效率和开销优化都需要

进一步提高^[36]。

5 结束语

人工智能技术已经应用于广域网端到端低时延网络传输中，并在传输层、应用层取得了一定的成果。本文分析网络传输时延的组成，并介绍了传输层中拥塞控制、应用层流媒体传输与机器学习技术相结合的研究。近年来，网络传输技术、机器学习模型技术的进一步发展，为网络智能传输带来更多的可能。我们认为，网络传输模型的泛化能力、大规模部署的开销、多目标学习之间的权衡、数据质量，是网络智能传输技术的新研究主题与挑战。

参考文献

- [1] LETAIEF K B, CHEN W, SHI Y M, et al. The roadmap to 6G: AI empowered wireless networks [J]. IEEE communications magazine, 2019, 57(8): 84–90. DOI: 10.1109/MCOM.2019.1900271
- [2] BRAUD T, LEE L H, ALHILAL A, et al. DiOS—an extended reality operating system for the metaverse [J]. IEEE multimedia, 2023, 30(2): 70–80. DOI: 10.1109/mmul.2022.3211351
- [3] ZELAYA R I, SUSSMAN W, GUMMESON J, et al. LAVA: fine-grained 3D indoor wireless coverage for small IoT devices [C]//Proceedings of the 2021 ACM SIGCOMM 2021 Conference. ACM, 2021: 123–136. DOI: 10.1145/3452296.3472890
- [4] NI Y Z, ZHENG Z L, LIN X S, et al. CellFusion: multipath vehicle-to-cloud video streaming with network coding in the wild [C]//Proceedings of the ACM SIGCOMM 2023 Conference. ACM, 2023: 668–683. DOI: 10.1145/3603269.3604832
- [5] ELBAMBY M S, PERFECTO C, BENNIS M, et al. Toward low-latency and ultra-reliable virtual reality [J]. IEEE network, 2018, 32(2): 78–84. DOI: 10.1109/MNET.2018.1700268
- [6] WANG M W, CUI Y, WANG X, et al. Machine learning for networking: workflow, advances and opportunities [J]. IEEE network, 2018, 32(2): 92–99. DOI: 10.1109/MNET.2017.1700200
- [7] JORDAN M I, MITCHELL T M. Machine learning: trends, perspectives, and prospects [J]. Science, 2015, 349(6245): 255–260. DOI: 10.1126/science.aaa8415
- [8] VIOLA I, AMIRPOUR H, VEGA M T. IXR '22: 1st workshop on interactive eXtended reality [C]//Proceedings of the 30th ACM International Conference on Multimedia. ACM, 2022: 7412–7413. DOI: 10.1145/3503161.3554781
- [9] ZUO X T, WANG M W, XIAO T X, et al. Low-latency networking: architecture, techniques, and opportunities [J]. IEEE Internet computing, 2018, 22(5): 56–63. DOI: 10.1109/MIC.2018.053681363
- [10] AHMAD J, WARREN A. FPGA based deterministic latency image acquisition and processing system for automated driving systems [C]// Proceedings of 2018 IEEE International Symposium on Circuits and Systems (ISCAS). IEEE, 2018: 1–5. DOI: 10.1109/ISCAS.2018.8351472
- [11] SUN L Y, ZONG T Y, WANG S Q, et al. Towards optimal low-latency live video streaming [J]. IEEE/ACM transactions on networking, 2021, 29(5): 2327–2338. DOI: 10.1109/TNET.2021.3087625
- [12] PEDERSEN H A, DEY S. Enhancing mobile video capacity and quality using rate adaptation, RAN caching and processing [J]. IEEE/ACM transactions on networking, 2016, 24(2): 996–1010. DOI: 10.1109/TNET.2015.2410298
- [13] XU Y D, ELAYOUBI S E, ALTMAN E, et al. Impact of flow-level dynamics on QoE of video streaming in wireless networks [C]//2013 Proceedings IEEE INFOCOM. IEEE, 2013: 2715–2723. DOI: 10.1109/INFCOM.2013.6565080
- [14] LANGLEY A, RIDDOCH A, WILK A, et al. The QUIC transport protocol: design and Internet-scale deployment [C]//Proceedings of the Conference of the ACM Special Interest Group on Data Communication. ACM, 2017: 183–196. DOI: 10.1145/3098822.3098842
- [15] IETF. QUIC: a UDP-based multiplexed and secure transport: RFC 9000 [S]. 2021
- [16] CUI Y, LI T X, LIU C, et al. Innovating transport with QUIC: design approaches and research challenges [J]. IEEE Internet computing, 2017, 21(2): 72–76. DOI: 10.1109/MIC.2017.44
- [17] SHI H, CUI Y, QIAN F, et al. DTP: deadline-aware transport protocol [C]// Proceedings of the 3rd Asia-Pacific Workshop on Networking. ACM, 2019: 1–7. DOI: 10.1145/3343180.3343191
- [18] IETF. Media over QUIC [EB/OL]. (2022-09-12) [2023-09-15]. <https://datatracker.ietf.org/doc/charter-ietf-moq/>
- [19] IETF. MoQ relay for support of deadline-aware media transport [EB/OL]. (2023-09-11) [2023-09-15]. <https://datatracker.ietf.org/doc/draft-moq-relay-for-deadline/>
- [20] BOVY C J, MERTODIMEDJO H T, HOOGHEMSTRA G, et al. Analysis of end-to-end delay measurements in Internet [EB/OL]. [2023-09-15]. https://www.researchgate.net/publication/233863840_Analysis_of_end-to-end_delay_measurements_in_Internet
- [21] MISHRA A, SHIN M, ARBAUGH W. An empirical analysis of the IEEE 802.11 MAC layer handoff process [J]. ACM SIGCOMM computer communication review, 2003, 33(2): 93–102. DOI: 10.1145/956981.956990
- [22] PHEMIUS K, MATHIEU B. Openflow: why latency does matter [C]//2013 IFIP/IEEE International Symposium on Integrated Network Management (IM 2013). IEEE, 2013: 680–683
- [23] CARDWELL N, CHENG Y C, GUNN C S, et al. BBR: Congestion-based congestion control: measuring bottleneck bandwidth and round-trip propagation time [EB/OL]. [2023-09-15]. <https://doi.org/10.1145/3012426.3022184>
- [24] WOLFMAN A, VOELKER G, THEKKATH C A. Latency analysis of TCP on an ATM network [EB/OL]. [2023-09-15]. <https://cseweb.ucsd.edu/~voelker/pubs/tcp-useenix94.pdf>
- [25] NOOR-A-RAHIM M, LIU Z L, LEE H, et al. 6G for vehicle-to-everything (V2X) communications: enabling technologies, challenges, and opportunities [J]. Proceedings of the IEEE, 2022, 110(6): 712–734. DOI: 10.1109/jproc.2022.3173031
- [26] SWAMINATHAN V, WEI S. Low latency live video streaming using HTTP chunked encoding [C]//Proceedings of 2011 IEEE 13th International Workshop on Multimedia Signal Processing. IEEE, 2011: 1–6. DOI: 10.1109/MMSP.2011.6093825
- [27] WINSTEIN K, BALAKRISHNAN H. TCP ex machina: computer-generated congestion control [C]//Proceedings of the ACM SIGCOMM 2013 conference on SIGCOMM. ACM, 2013: 123–134. DOI: 10.1145/2486001.2486020
- [28] SIVARAMAN A, WINSTEIN K, THAKER P, et al. An experimental study of the learnability of congestion control [J]. ACM SIGCOMM computer communication review, 2014, 44(4): 479–490
- [29] JAY N, NOGA R, BRIGHTEN G, et al. A deep reinforcement learning perspective on internet congestion control [EB/OL]. (2019-05-21) [2023-09-15]. <https://arxiv.org/pdf/1810.03259.pdf>
- [30] HUANG T C, ZHOU C, JIA L C, et al. Learned Internet congestion control for short video uploading [C]//Proceedings of the 30th ACM International Conference on Multimedia. ACM, 2022: 3064–3075. DOI: 10.1145/3503161.3548436
- [31] DONG M, MENG T, ZARCHY D, et al. PCC vivace: online-learning congestion control [C]//15th USENIX Symposium on Networked Systems Design and Implementation (NSDI 18). ACM, 2018: 343–356
- [32] ZHANG L, CUI Y, PAN J C, et al. Deadline-aware transmission control for real-time video streaming [C]//Proceedings of 2021 IEEE 29th International Conference on Network Protocols (ICNP). IEEE, 2021: 1–6. DOI: 10.1109/ICNP52444.2021.9651971
- [33] NIE X H, ZHAO Y J, LI Z H, et al. Dynamic TCP initial windows and congestion control schemes through reinforcement learning [J]. IEEE journal on selected areas in communications, 2019, 37(6): 1231–1247. DOI: 10.1109/JSAC.2019.2904350
- [34] ABBASLOO S, YEN C Y, CHAO H J. Classic meets modern: a pragmatic learning-based congestion control for the Internet [C]//Proceedings of the Annual conference of the ACM Special Interest Group on Data Communication on the applications, technologies, architectures, and protocols for computer communication. ACM, 2020: 632–647. DOI: 10.1145/3387514.3405892
- [35] ZHANG L, ZHU K W, PAN J C, et al. Reinforcement learning based congestion control in a real environment [C]//Proceedings of 2020 29th International Conference on Computer Communications and Networks (ICCCN). IEEE, 2020: 1–9. DOI: 10.1109/ICCCN49398.2020.9209750

- [36] ZHANG L, CUI Y, WANG M W, et al. DeepCC: bridging the gap between congestion control and applications via multiobjective optimization [J]. IEEE/ACM transactions on networking, 2022, 30(5): 2274–2288. DOI: 10.1109/TNET.2022.3167713
- [37] MAO H Z, NETRAVALI R, ALIZADEH M. Neural adaptive video streaming with pensieve [C]//Proceedings of the Conference of the ACM Special Interest Group on Data Communication. ACM, 2017: 197 – 210. DOI: 10.1145/3098822.3098843
- [38] HUANG T C, ZHANG R X, ZHOU C, et al. QARC: video quality aware rate control for real-time video streaming based on deep reinforcement learning [C]//Proceedings of the 26th ACM International Conference on Multimedia. ACM, 2018: 1208 – 1216. DOI: 10.1145/3240508.3240545
- [39] ROBITZA W, GARCIA M N, RAAKE A. A modular HTTP adaptive streaming QoE model—candidate for ITU-T P.1203 (“P. NATS”) [C]//Proceedings of 2017 Ninth International Conference on Quality of Multimedia Experience (QoMEX). IEEE, 2017: 1–6. DOI: 10.1109/QoMEX.2017.7965689
- [40] ESWARA N, ASHIQUE S, PANCHBHAI A, et al. Streaming video QoE modeling and prediction: a long short-term memory approach [J]. IEEE transactions on circuits and systems for video technology, 2020, 30(3): 661–673. DOI: 10.1109/tcsvt.2019.2895223
- [41] LAI Z Q, HU Y C, CUI Y, et al. Furion: engineering high-quality immersive virtual reality on today’s mobile devices [C]//Proceedings of the 23rd Annual International Conference on Mobile Computing and Networking. ACM, 2017: 409 – 421. DOI: 10.1145/3117811.3117815
- [42] ZUO X T, YANG J Y, WANG M W, et al. Adaptive bitrate with user-level QoE preference for video streaming [C]//Proceedings of IEEE INFOCOM 2022 – IEEE Conference on Computer Communications. IEEE, 2022: 1279–1288. DOI: 10.1109/INFOCOM48880.2022.9796953
- [43] LV G R, WU Q H, WANG W R, et al. Lumos: towards better video streaming QoE through accurate throughput prediction [C]//Proceedings of IEEE INFOCOM 2022 – IEEE Conference on Computer Communications. IEEE, 2022: 650–659. DOI: 10.1109/INFOCOM48880.2022.9796948
- [44] ZHANG J, SHI H, CUI Y, et al. To punctuality and beyond: meeting application deadlines with DTP [C]//Proceedings of 2022 IEEE 30th International Conference on Network Protocols (ICNP). IEEE, 2022: 1–11. DOI: 10.1109/ICNP55882.2022.9940391
- [45] CUI Y, WANG L, WANG X, et al. FMTCP: a fountain code-based multipath transmission control protocol [J]. IEEE/ACM transactions on networking, 2015, 23(2): 465–478. DOI: 10.1109/TNET.2014.2300140
- [46] ZHENG Z L, MA Y F, LIU Y M, et al. XLINK: QoE-driven multi-path QUIC transport in large-scale video services [C]//Proceedings of the 2021 ACM SIGCOMM 2021 Conference. ACM, 2021: 418–432. DOI: 10.1145/3452296.3472893
- [47] ZUO X T, CUI Y, WANG X, et al. Deadline-aware multipath transmission for streaming blocks [C]//Proceedings of IEEE INFOCOM 2022 – IEEE Conference on Computer Communications. IEEE, 2022: 2178–2187. DOI: 10.1109/INFOCOM48880.2022.9796942
- [48] LIU X, ZHANG F J, HOU Z Y, et al. Self-supervised learning: generative or contrastive [J]. IEEE transactions on knowledge and data engineering, 2023, 35(1): 857–876. DOI: 10.1109/TKDE.2021.3090866
- [49] BROWN T, MANN B, RYDER N, et al. Language models are few-shot learners [EB/OL]. (2020-05-28) [2023-07-25]. <https://arxiv.org/abs/2005.14165>
- [50] HU S H, ZENG G X, BAI W, et al. Aeolus: a building block for proactive transport in datacenter networks [J]. IEEE/ACM transactions on networking, 2022, 30(2): 542–556. DOI: 10.1109/TNET.2021.3119986
- [51] BAI W, HU S H, CHEN K, et al. One more config is enough: saving (DC) TCP for high-speed extremely shallow-buffered datacenters [J]. IEEE/ACM transactions on networking, 2021, 29(2): 489–502. DOI: 10.1109/TNET.2020.3032999
- [52] WU H T, FENG Z Q, GUO C X, et al. ICTCP: Incast Congestion Control for TCP in data center networks [C]//Proceedings of the 6th International Conference. ACM, 2010: 1 – 12. DOI: 10.1145/1921168.1921186

作者简介



廖乙鑫, 北京邮电大学在读硕士研究生; 主要研究领域为低时延网络传输、视频传输。



王子逸, 北京邮电大学副研究员; 主要研究领域为低时延网络传输、流媒体传输、边缘计算等; 在《IEEE/ACM Transactions on Networking》、INFOCOM等权威期刊与会议上发表论文多篇。



崔勇, 清华大学长聘教授、网络技术研究所所长, 教育部“长江学者”特聘教授, 首届“青年长江学者”获得者, 中国互联网协会学术工作委员会秘书长, 中国通信学会边缘计算委员会副主任委员, 《IEEE Transactions on Parallel and Distributed Systems》等4个IEEE期刊的编委, 曾长期担任国际标准工作组主席; 主要研究方向为低时延传输技术、视频分析、内容安全、流媒体传输、网络数字孪生、网络AI等; 获国家优秀青年科学基金和教育部新世纪人才等项目持续支持, 2019年成功在北京组织国际网络通信领域顶级会议 Sigcomm’19 并担任大会副主席, 参与研制我国第一台“IPv6核心路由器”, 参与建设中国下一代互联网示范工程 CNGI-CERNET2; 获国家科技进步二等奖1次、国家技术发明二等奖1次, 多次获得国家信息产业重大技术发明; 发表论文100余篇, 获国家发明专利40余项, 完成RFC国际标准10余项, 出版学术著作4部。

NetGPT:超越个性化生成服务的内生智能网络架构



NetGPT: An AI-Native Network Architecture for Provisioning Beyond Personalized Generative Services

陈宇轩/CHEN Yuxuan, 李荣鹏/LI Rongpeng,
张宏纲/ZHANG Honggang
(浙江大学, 中国 杭州 310007)
(Zhejiang University, Hangzhou 310007, China)

DOI: 10.12142/ZTETJ.202305011

网络出版地址: <http://kns.cnki.net/kcms/detail/34.1228.TN.20231016.1554.016.html>

网络出版日期: 2023-10-17

收稿日期: 2023-08-02

摘要: 提出了基于边缘和云端部署相匹配大型语言模型（LLM）的内生智能网络架构 NetGPT 方案。边缘 LLM 可以有效地利用基于位置的信息进行个性化的补充，从而与云端 LLM 进行有效交互。通过在边缘和云端部署开源 LLM，验证了 NetGPT 的可行性。认为面向 NetGPT 的内生智能网络架构的工作重点是通信和计算资源的深度集成以及 AI 逻辑工作流的灵活设计。认为 NetGPT 是一种可提供个性化的生成式服务的、有前途的内生智能网络架构。

关键词: LLM; 内生智能网络架构; 云边协同; 个性化生成服务

Abstract: The NetGPT framework, which is founded upon the alignment of large language models (LLMs) tailored for both edge and cloud deployments, is introduced. Edge-oriented LLMs harness location-based data to effectively personalize content augmentation, facilitating seamless interactions with their cloud-based counterparts. The viability of the NetGPT paradigm is empirically substantiated through the deployment of open-source LLMs at both the edge and cloud strata. It is believed that within the realm of endogenous intelligent network architectures designed to support NetGPT, the central emphasis rests on the profound integration of communication and computational resources, coupled with the adaptability in the design of AI logic workflows.

Keywords: LLM; AI-native network architecture; edge-cloud collaboration; personalized generative services

引用格式: 陈宇轩, 李荣鹏, 张宏纲 . NetGPT: 超越个性化生成服务内生智能网络架构 [J]. 中兴通讯技术, 2023, 29(5): 68–75. DOI: 10.12142/ZTETJ.202305011

Citation: CHEN Y X, LI R P, ZHANG H G. NetGPT: an AI-native network architecture for provisioning beyond personalized generative services [J]. ZTE technology journal, 2023, 29(5): 68–75. DOI: 10.12142/ZTETJ.202305011

随着深度学习从 AlphaGo 到 ChatGPT 应用的转变，人工智能（AI）的作用将在 6G 网络中不断体现。一方面，边缘计算能力的提升将会使网络资源得到有效安排，服务质量（QoS）得到改善，以 AI 为高效的高效服务供给研究也受到普遍重视。另一方面，一个 AI 智能模型的应用往往局限于某些场景或任务。例如，大型语言模型（LLMs）^[1]在各种自然语言处理（NLP）和计算机视觉任务中表现出色，而在实际场景中，要使预训练 LLM 遵循人类意图生成个性化输

出，则需要对 LLM 进行微调^[2]。仅在集中式云端部署 LLM 以寻求模型的个性化微调将为云端带来多个完整模型参数副本，因此是一种低效的做法。

为了改善 LLM 的个性化问题，找到合适的云边协同方法至关重要^[3]。与仅在云端部署 LLM 的方法相比，用云边协同的方式来部署大模型有多重优点。这种方法能赋予边缘服务器较大的自由度，从而可以部署多种微调的 LLM，适应环境差异以实现服务个性化、定制化。同时，这种云边协同可以将数据丰富的生成式设备连接到更多邻近服务器上，从而减少向更多远程云服务器上传数据的延迟，并节省通信开销。把生成式 LLM 融入边缘网络有望促进通信和计算（C&C）资源的高效使用。

基金项目：国家自然科学基金项目（62071425）；浙江省“领雁”计划项目（2022C01093）；浙江省杰出青年基金项目（LR23F010005）

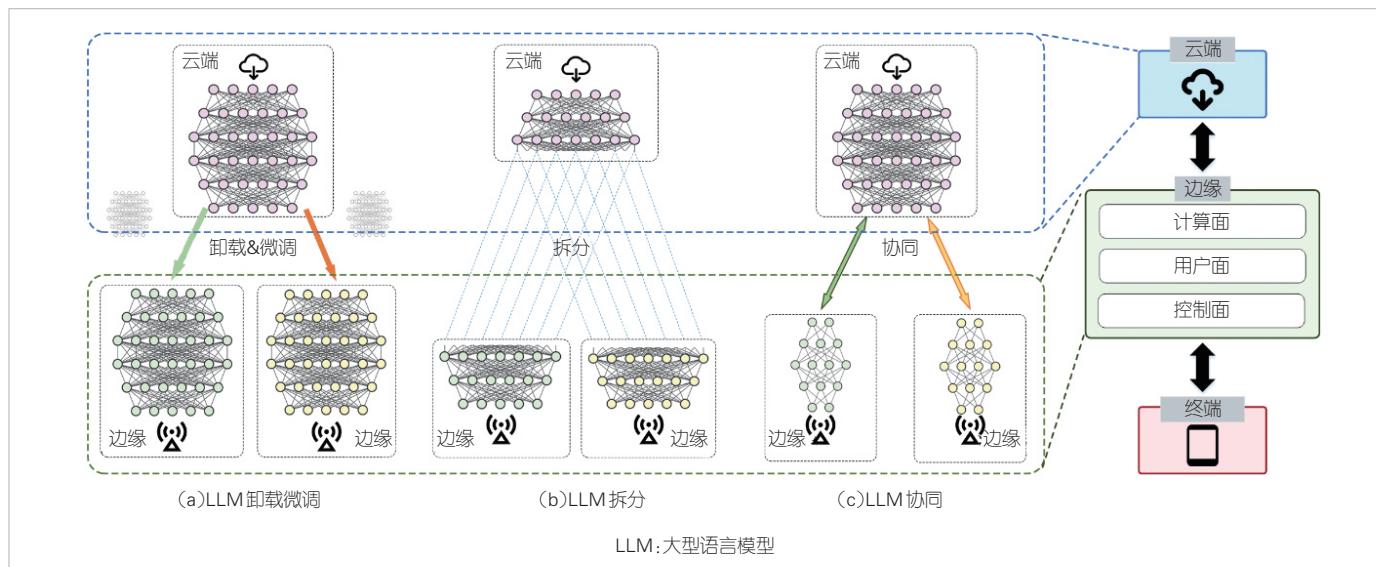
如图1所示，有几种方法可以实现LLM的云边协同部署，如本地微调、模型拆分等。具体而言，本地边缘服务器可通过卸载云端训练的LLM来定制LLM，实现个性化、定制化服务，满足用户喜好及场景需求。在此场景下，联邦学习或并行训练可作为一种辅助手段来实现调优^[2, 4]。但对完整LLM的重复微调意味着庞大的计算，并有可能会给模型开发人员带来知识产权上的困扰，因此该方法在实际应用中也存在着诸多问题。同时，对边缘处的LLM整体进行强制拟合可能会使边缘服务器受限，计算资源紧张，从而导致边缘计算所需开销过于庞大。另外，卸载LLM也会产生明显的通信开销。另一种方案是将LLM拆分后部署到云端和边缘服务器，通过在边缘部署一些大规模深度神经网络(DNN)层，将剩余的层留给云端，从而有效平衡边缘服务器和云端服务器间的计算资源。在模型划分中，如何有效地将DNNs从边缘到云端进行划分是最具挑战性的问题之一，这需要在最小化端到端时延的同时，为边缘服务器保留足够小的模型尺寸^[4]。考虑到典型的LLM中有数十亿个参数，这样的模型划分可能会非常复杂，LLM中广泛采用的参差链接也可能会限制合适划分点的选择。另外，LLM可能会泄露训练数据中的隐私细节^[5]，因此直接以局部微调与模型分割的方法来实现云边协同，也会存在一定的挑战。

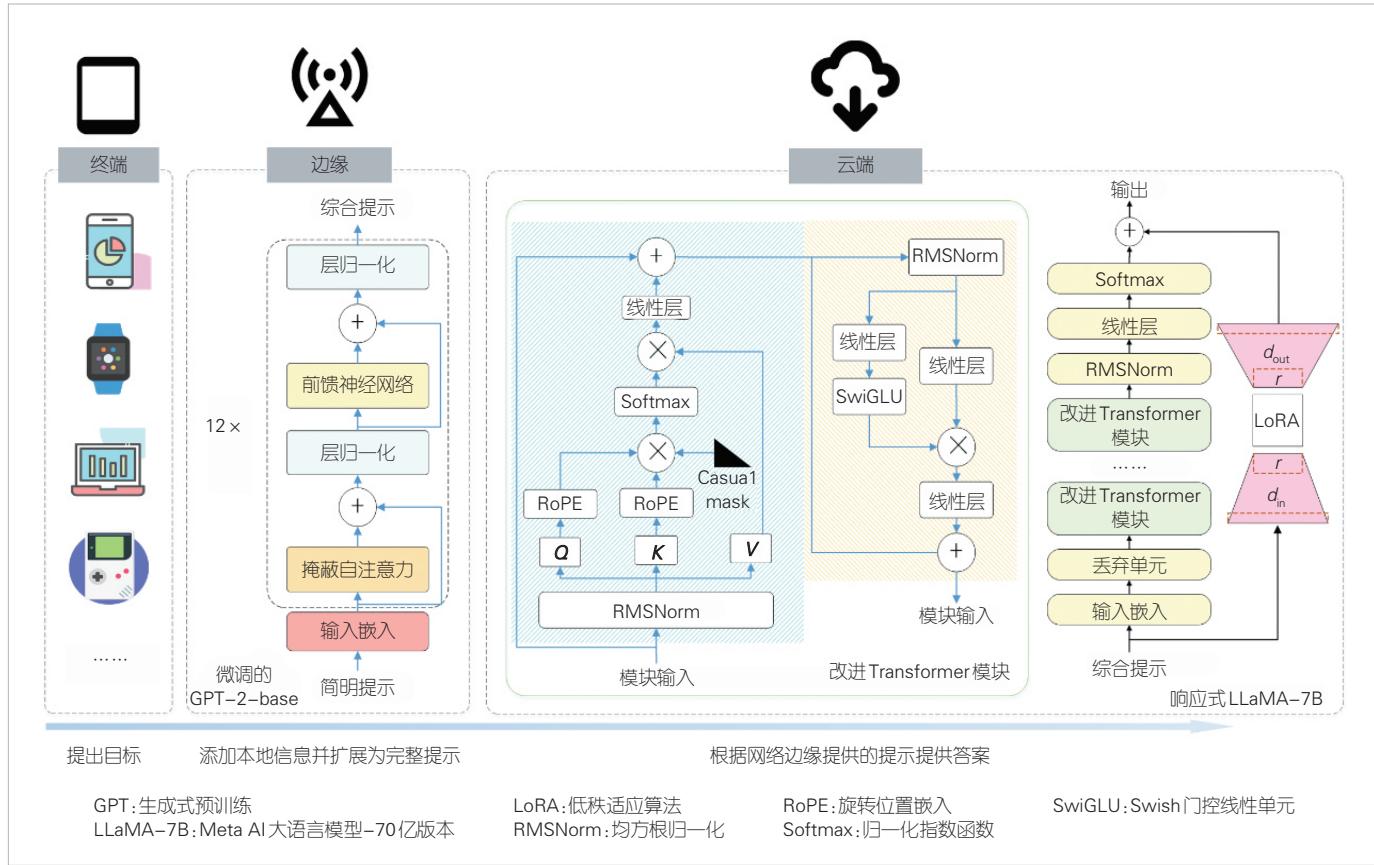
本文中，我们提出内生智能网络架构NetGPT，基于云边不均衡的资源分布，实现了边缘与云端之间不同尺寸功能性LLM的协同。与具有解耦C&C资源的AI外生网络明显不同，NetGPT能够使用融合C&C对边缘部署更小的LLM，对云端部署更大的LLM，以进行有目的的云边协同计算，从而

提供个性化的内容生成服务。此外，NetGPT还集成并发展了有逻辑的AI工作流，以识别具有相同性能的通信链路。例如，在NetGPT中，假设边缘LLM提供满意的内容，那么性能驱动的通信链路会在边缘终止以加速响应。否则，在即时学习^[6]理念的影响下，边缘LLM能够推断上下文并主动附加(或填充)部分个性化信息，从而在云端得到更加全面的效果。同时，边缘LLM有助于为智能网络管理和调度(例如用户意图推断、流行度预测等)提供统一的解决方案。因此，NetGPT符合C&C深度融合的发展趋势，代表了一种由LLM驱动的内生智能网络架构。

1 NetGPT的实现

我们提出了一个云边协同框架，具体如图2所示。通过在云端和边缘(如基站)使用不同的预训练LLM，可以完成个性化的生成服务。受限于开源LLM的可用性，我们选择并部署了LLaMA-7B模型(Meta AI大语言模型-70亿版本)^[7]和基于GPT-2的模型。这两个模型分别由大约67亿个和1亿个参数组成，并部署在云端和边缘。值得注意的是，NetGPT可以根据需要使用其他LLM。基于此，我们对云边LLM协同NetGPT的实现细节进行了深入研究。首先，我们对作为LLM基础的Transformer进行了总体概述，并列出了LLaMA-7B模型和GPT2基础模型两个LLM的详细DNN结构。随后，我们探讨了在计算受限的设备上微调LLM的有效途径，并且展示了其对基于位置的个性化生成服务的效果。





▲图2 面向NetGPT的云边协同计算框架

1.1 Transformer概述

Transformer 已被广泛用作 LLM 中多层解码器的基础模型。Transformer 是通过使用多层自注意力和前馈神经网络 (FNN) 来构建 DNN 结构。自注意力依赖于由 query、key 和 value 矩阵 (即 Q 、 K 和 V) 定义的参数化注意力头，通过推导不同的权重并将其分配给序列中的不同位置来计算输入序列内的内部相关性。在 FNN 中，每个位置的表示使用的是非线性变换。此外，Transformer 采用层归一化等技术来缓解梯度消失的问题。

1.2 边缘和云端 LLM 的 DNN 结构

1.2.1 GPT-2-base 模型的 DNN 结构

GPT-2-base 模型是 GPT-2 系列中最小参数的版本，包含了原 Transformer 结构的 12 层堆叠 (即 8 头自注意力子层和 FNN 子层)。该模型利用正余弦位置的固定绝对位置编码方式来预变换输入序列。此外，GPT-2 使用修正线性单元 (ReLU) 来激活函数。该模型具有相对优异的性能和较低的计算要求，因此适合部署在网络边缘。

1.2.2 LLaMA 模型的 DNN 结构

LLaMA 经过大量无标签数据的训练，非常适合下游任务的微调，同时也有多种参数版本^[7]。与 GPT-3 相比，LLaMA 结合了多项特定增强功能，从而在保持相似性能的同时显著减少了参数数量^[7]。为了提高训练稳定性，LLaMA 采用了对各子层的输入进行归一化而非对输出进行归一化的方式。此外，LLaMA 使用一种简化的替代方案，即采用均方根层归一化 (RMSNorm) 函数^[8]，利用均方根而非标准差进行归一化处理。此外，RMSNorm 引入了可学习的缩放因子进行自适应特征缩放，从而增强具有不同值域的各种特征的归一化效果。其次，LLaMA 用 Swish-Gated 线性单元 (SwiGLU)^[9]取代了 ReLU 激活函数。该激活函数将 Swish 函数 (即 $f_{\text{Swish}}(x) = x \cdot \sigma(\beta x)$ ，其中 $\sigma(x) = \frac{1}{1 + e^{-x}}$ ， β 为可训练参数) 和门控线性单元 (GLU) (即 $f_{\text{GLU}}(x) = x \cdot \sigma(Wx + b)$ ， W 和 b 为可训练参数) 相结合，从而可以根据输入以更有选择性的方式激活神经元，在输入发生改变时更加平滑，以有效捕获复杂的非线性关系。最后，LLaMA 引入了旋转位置嵌入 (RoPE)^[10]，利用预设的旋转矩阵对位置信息进行编

码，自然地将显式相对位置依赖性纳入自注意力公式中。相较于给序列中每个位置分配独特编码表示的句对位置编码方式，RoPE中所采用的相对位置编码方式能更有效地对上下文信息中的远程依赖关系进行建模，易于直观理解，并且在实验中表现出优秀的性能。

1.3 微调技术

1.3.1 基于低秩自适应的轻量级微调

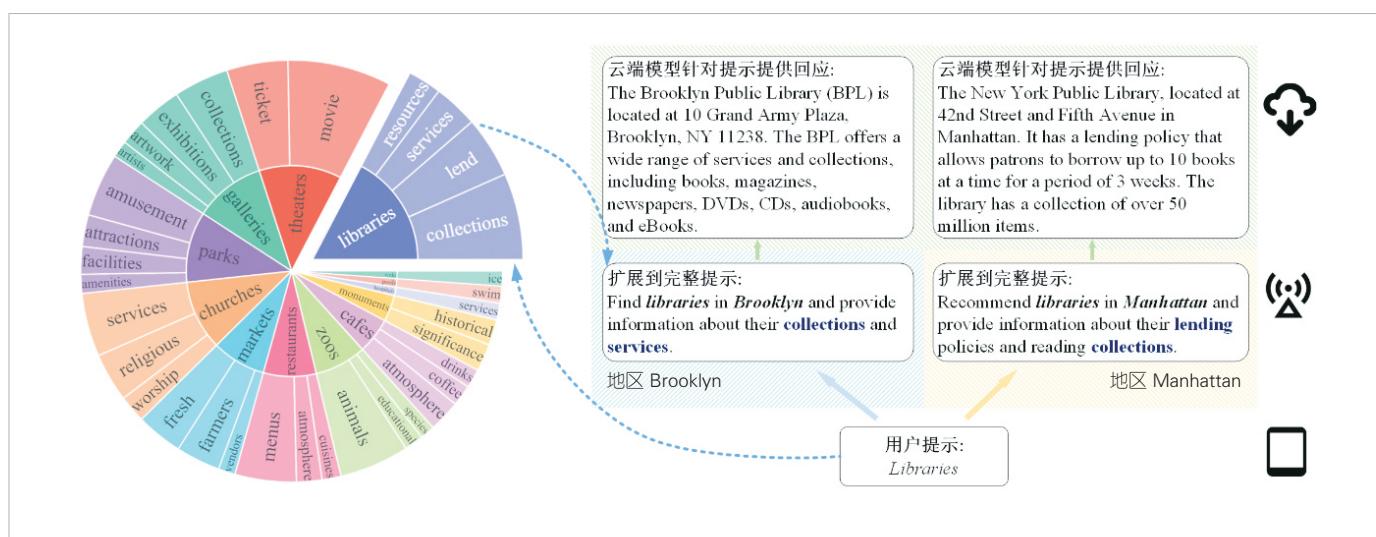
LLaMA 缺乏生成响应式文本的能力^[7]，因此仍需进行微调，但直接微调会耗费大量的计算资源。例如，直接微调 LLaMA-7B 模型需要 112 GB 视频随机存取存储器（VRAM），这超过 NVIDIA A100 Tensor Core GPU 的容量。因此，我们采用低秩适应（LoRA）技术^[11]在消费级硬件上进行参数高效的微调。为了微调完整的参数矩阵 $\mathbf{W} \in \mathbb{R}^{d_{\text{in}} \times d_{\text{out}}}$ ，LoRA 添加了一条旁路路径，通过使用两个具有内在秩 r 的下游参数矩阵 $\mathbf{A} \in \mathbb{R}^{d_{\text{in}} \times r}$ 和 $\mathbf{B} \in \mathbb{R}^{r \times d_{\text{out}}}$ 来模拟矩阵更新 $\Delta\mathbf{W}$ 。也就是说，在 $r \ll \min(d_{\text{in}}, d_{\text{out}})$ 的条件下，LoRA 成功地将大型参数矩阵 $\Delta\mathbf{W}$ 转换为 $\Delta\mathbf{W} \approx \mathbf{AB}$ 的低秩矩阵。实验证明，微调后的 LLaMA-7B 模型只消耗了 28 GB VRAM 而没有明显增加训练时间。另外，微调后需要的存储空间由 12.55 GB 明显缩小至 50 MB 左右¹。基于 LoRA，我们使用 Stanford Alpaca 数据集^[12]对 LLaMA-7B 模型进行微调，成功得到响应式 LLaMA-7B 模型。

1.3.2 在 LLM 引导下收集数据

为了实现个性化边缘 LLM，GPT-2-base 模型需要通过附加基于位置的额外信息对“简明提示”进行扩展。本质上，基于 5G 接入与移动管理功能（AMF）所储存的附属 BS 位置能够很容易得到定位信息。同时，为了补充更全面的信息，本文中我们用 self-instruct 方法^[13]，即使使用 OpenAI 的 TextDavinci-003 模型来生成有效的文本样本。在每个地区，我们使用一组手动编写的位置相关提示与 OpenAI 的 Text-Davinci-003 模型进行交互，并将生成响应文本作为“综合提示”。与此相对应，可以收集一系列“简明提示”和“综合提示”之间的映射关系。考虑到边缘 LLM 的规模和任务复杂性，我们收集了一个包含大约 4 000 个样本的数据集，将 GPT-2 的模型微调为即时完成模型。值得注意的是，针对高通用性的场景，可以使用更大规模的 LLM 来增强边缘 LLM，也可以采用 LoRA 等微调技术。

1.4 性能展示

图 3 显示的是 NetGPT 的性能。其中，边缘 LLM 能够根据左侧的图表补充“简明提示”，从而产生每一个对应的“综合提示”。不同基站也加入基于位置的个性化信息来满足不同需求。随后，边缘 LLM 对用户提出的“简明提示”进行处理，并向云端反馈补充提示信息，之后云端 LLM 再给出完整的回应。从图 3 右侧可看出，NetGPT 能产生基于位置的不同回应，这反映出通过对云边进行高效的协同处理可以实现个性化的服务生成。



▲图 3 NetGPT 基于位置的个性化生成服务的性能展示

¹此类统计数据是在 $r=8$ 且与学习率相关的标量因子等于 16 的配置下获得的。

2 面向NetGPT的内生智能网络架构

我们认为，NetGPT为蜂窝网络向内生智能网络架构的转变提供了机会，给用户带来个性化、网络化以及包容性智能，并赋予用户更多权限，以随时随地访问生成服务。尽管如此，这一转变是需要付出代价的。这需要对网络架构做出实质上的改变，不仅仅是在边缘位置安装服务器机架，以及在本地中断流量以进行边缘处理。相对于传统面向连接通信系统，典型服务在两个具体终端间建立了连接，通信源与目的地是最终用户清晰界定的。NetGPT需通过更加隐性的方式来建立产生性能驱动的连接。此外，NetGPT涉及更频繁的数据收集和处理模块，因此计算资源应保持一致。如图4所示，NetGPT需要在无线接入网络（RANs）中设计深度融合的C&C。在这些创新功能之外，NetGPT还需要设计逻辑AI工作流来创建（超越）生成服务的编排。

2.1 RAN中的融合C&C

为了有效地组织同时覆盖地面通信和非地面通信的异构资源，用于NetGPT的RAN首先必须提供用于无缝连接控制的控制面（CP），以支持用户平面（UP）中的提示和生成信息传输。这些组件可按5G及5G的先进技术来开发。此外，引入一个独立的计算平面（CmP）来协调计算资源，并执行AI相关功能以促进生成服务的部署和更新也是非常必要的。

2.2 数据处理和隐私保护

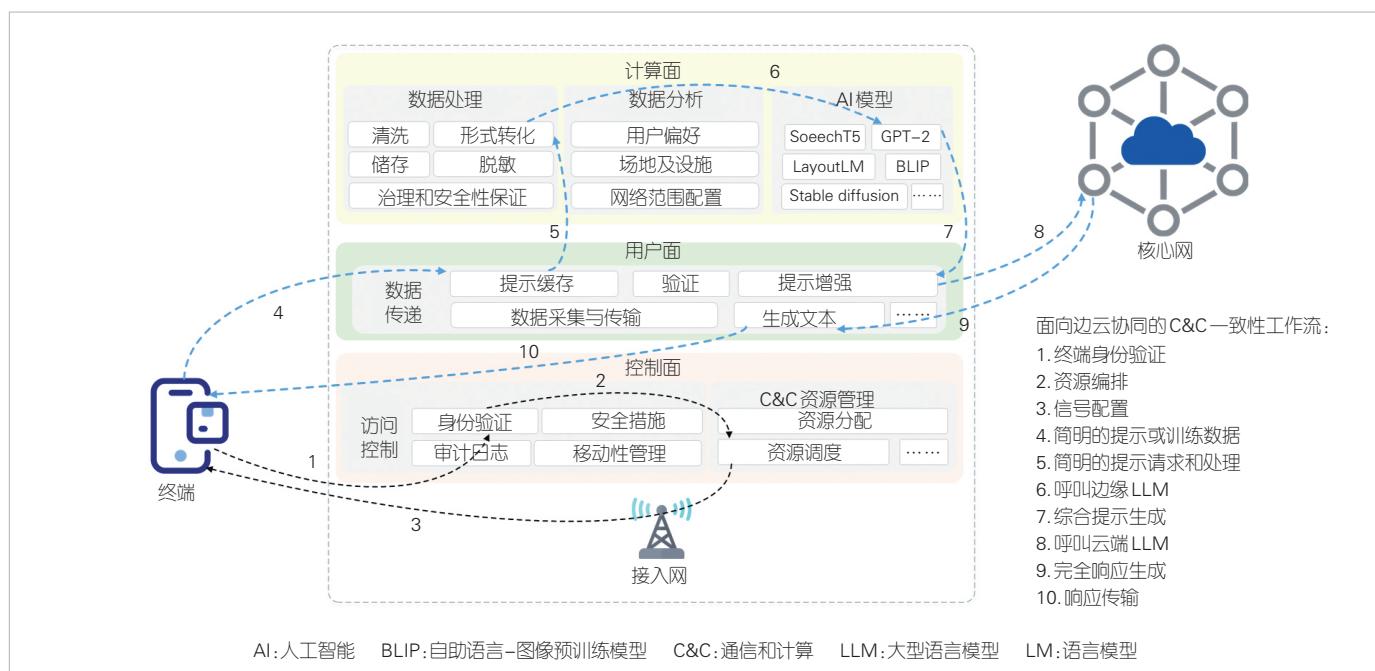
如1.2节所述，数据处理（如数据收集和微调）技术的广泛使用为LLM的产生提供了依据。在采集并储存数据的同时，引入数据脱敏模块也非常重要，这样可以规避隐私风险并对嵌入数据进行隐私防护。同时，数据策略执行模块将默认执行，并根据监管和非监管规则（如地理限制）处理数据，以确保数据处理的完整性和合法性。另外，基于监管与数据使用政策，可设计若干数据处理模型库，并设置适当的访问权限控制，使得有需要的实体能够调用这些数据服务。

2.3 个性化分析

为了创建高度自定义NetGPT，应增强对面向位置相关信息的分析，以支持个性化生成AI服务的定义和操作。例如，一方面，可以专门收集该地区的场地和设施信息来训练边缘LLM；另一方面，用户服务节点（USN）也可以包含最终用户级别的定制服务，以满足多样化的客户需求，同时也进一步支持用户档案的建立和所联系终端的表征。

2.4 C&C资源管理

NetGPT资源编排是未来蜂窝网络供应服务的组成部分，它与其他网络服务资源编排具有一定的相似性，包括建立连接和计算资源分配。但由于其涉及的资源范围跨越云到终端等分布式节点，这也带来了一些挑战。因此，需要在无线控制信令或无线数据协议上携带新的协议栈来传输AI生成消



▲图4 面向NetGPT的内生智能网络架构和逻辑AI工作流

息并实现模型的更新和分发。

2.5 逻辑AI工作流

为了有效地提供AI服务，需要首先制定部分逻辑AI工作流以解析并安排NetGPT服务。值得注意的是，逻辑AI工作流能够协调分散在边缘与云端的一系列网络功能，连续地提供“简明提示”“综合提示”和“生成响应”，规范数据处理与分析，从而在CmP中进行个性化LLM训练。另外，还需要在服务部署过程中将逻辑AI工作流映射为物理资源，来满足有关服务QoS需求。由于工作流程涉及许多网络功能，其处理过程可以建立在串行或有向无环图上。另一方面，逻辑AI工作流并不局限于产生服务。通过更为可定制的途径，逻辑AI工作流极大地促进了QoS的改进。

3 基于LLM的网络管理与协调统一解决方案

除了提供个性化的生成服务外，面向NetGPT的内生智能网络架构还可以在边缘LLM上为智能网络管理和编排提供统一的解决方案。

3.1 流行度预测

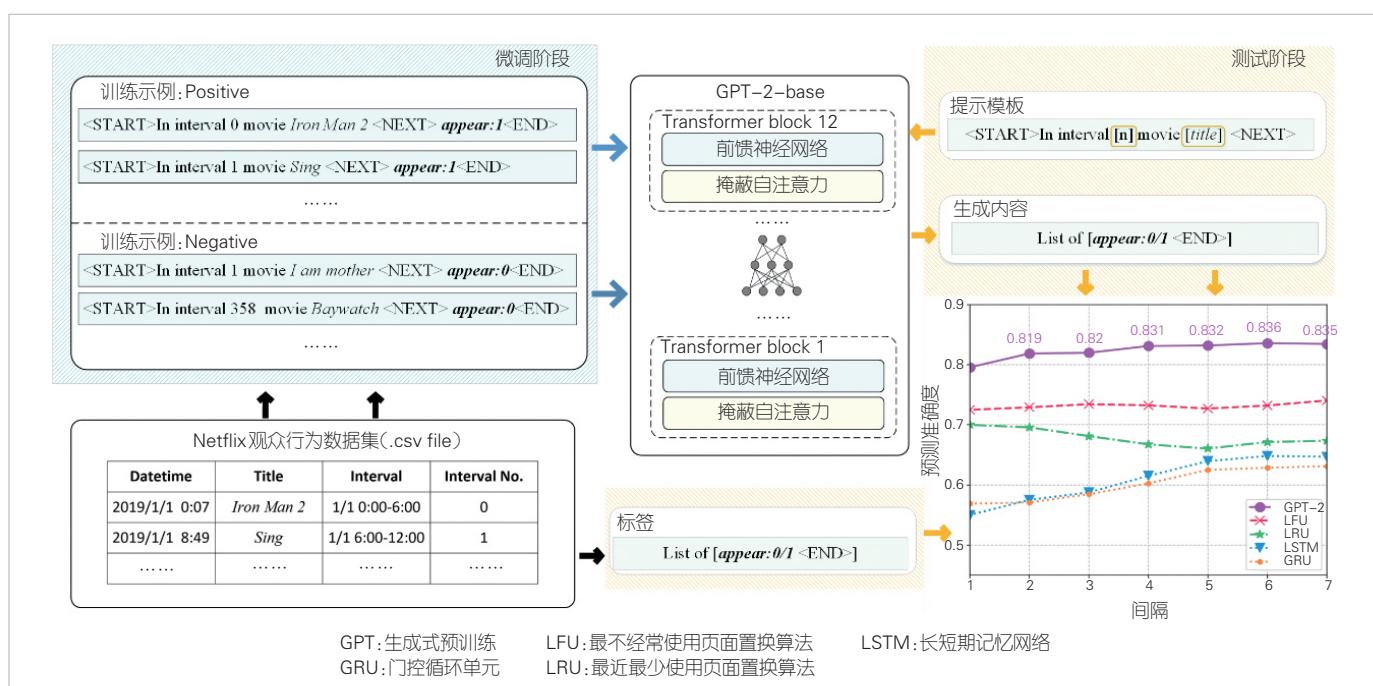
通过让C&C资源适应预测的需求，流行度预测可以显著提高网络效率^[14]。鉴于DNN结构基本原理，GPT-2能够从RAN的附属终端历史访问记录中解释用户的偏好。另外，

通过结合特定地点的数据，边缘LLM可以更好地捕捉每个地区独有的个性化特征。

为了证明边缘LLM（即GPT-2-base模型）的预测能力，我们以Netflix观众行为数据集为例进行了实验。结果表明，该方法能够有效地预测观众行为。为了解决数据稀疏性问题，我们先根据6 h循环对时间范围进行区间划分和编号标注；然后，根据每一部电影在特定区间内的存在情况，选取频率最高的20部电影进行标注。得益于CmP中的数据格式功能，相关的历史信息被转换为符合特定模板的一些自然语言。例如，“In interval 1, movie ‘Iron Man 2’ appear :1”表示电影《Iron Man 2》出现在区间1，与图5左下方所给的某一具体日期时间相对应。同时我们添加了特殊的标记来创建提示模板，从而帮助语言模型进行信息理解和生成响应。边缘LLM可在直接微调后根据提示模板的格式生成标签，即电影是否出现在间隔下。此外，为了增强模型的通用性，我们专门利用数据集中上半年的数据进行实验，并将95%的内容作为训练集，剩余的5%作为测试集。图5最后给出了边缘LLM的预测精度。可以观察到，针对这一任务，GPT-2显示了一个可以接受的精度水平，并明显优于其他经典算法。

3.2 意图推断

意图驱动网络旨在解决基于模板的服务在垂直业务中应



▲图5 流行度预测的边缘大型语言模型

用时难度增加的问题。意图驱动网络需要在取代人工配置网络和对网络问题做出反应之前，充分理解客户的实时需求^[15]。如何准确地理解客户的意图，并将其转化为可行的网络配置，是最根本的问题之一。边缘LLM能够满足这种意图识别过程^[15]，并精确理解口头陈述。通过采用类似于上文提到的微调方法，边缘LLM可以理解和提取任意自然语言输入中的关键词。值得注意的是，这样的微调可以很容易地完成。在实验中，我们仅使用了4 000个输入关键字对即可实现优秀的效果。图6展示了GPT-2-base识别客户意图的相应能力。从图6可以看出，当一个用户想要建立从Access 1到Cloud 2的10 Gbit/s连接并提供流量保护时，GPT-2-base模型能够很方便地提取出所需的关键词，且不受语句差异的影响。因此，GPT-2-base模型避免了繁琐的模板设计和客户学习过程。相对于传统NLP工具，LLM显示了更加强大的意图驱动网络实现功能。

4 结束语

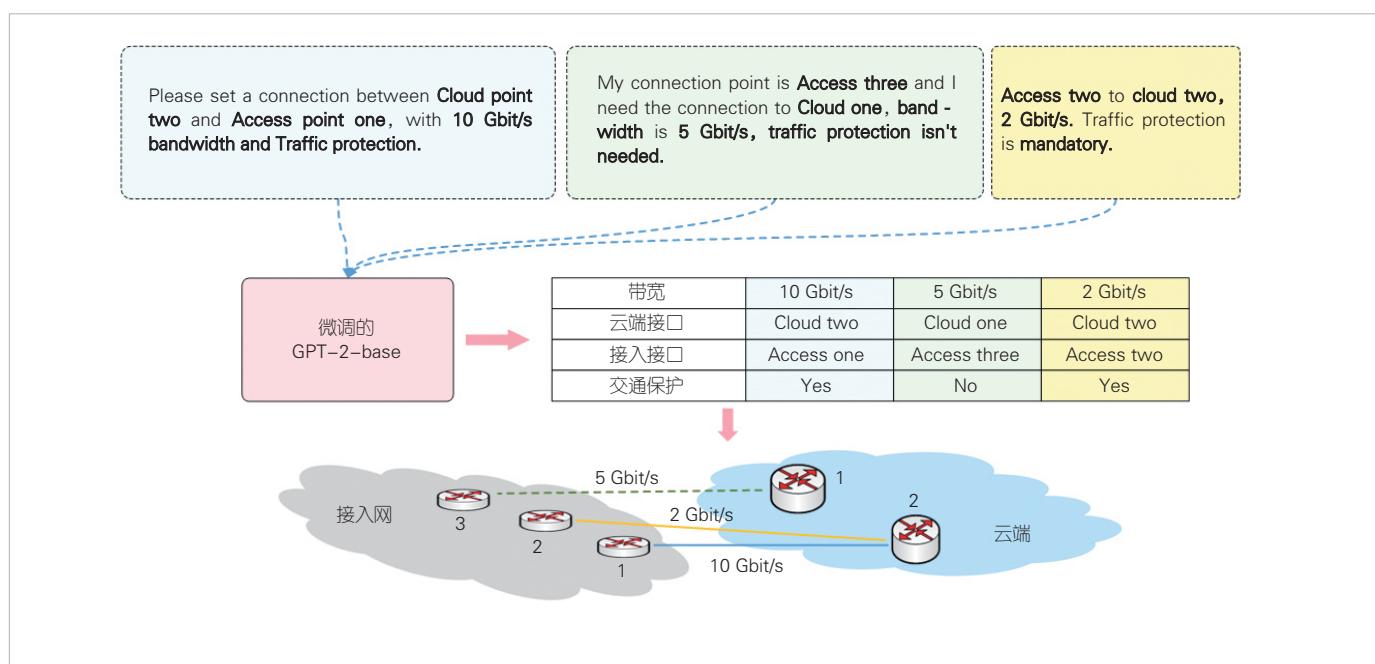
本文中，基于LLM，我们提出了一种内生智能网络架构NetGPT，从而提供超越个性化生成内容的网络服务。通过有效的云边协同，我们在边缘和云端部署一些具有代表性的开源LLM（如GPT-2-base模型和LLaMA模型）。在采用基于LoRA的参数高效微调技术后，我们再对开源LLM的一致

性进行评估。实验证明，NetGPT用于基于位置的个性化服务是可行的。除此之外，我们还强调了NetGPT所需的一些实质性架构更改（例如，深度C&C集成和逻辑AI工作流）。在此基础上，我们提出一个可能的边缘LLM许可网络管理与编排统一人工智能解决方案。

尽管NetGPT是一种很有前景的架构，但在本文中我们并未讨论其面临的研究挑战。要成功地部署NetGPT，必须解决如下几个问题：

- 如何在设备终端上实现推理和微调，以符合受限成本中计算能力的实质性约束？
- 如何模仿新必应²实现LLM在线学习来满足边缘无线环境动态性要求？
- 由于数值推理的灵敏度有限，可能存在幻觉效应，如何进一步提高LLM的严谨性，从最新的LLM中可以得到哪些启示？同时如何将LLM评估指标与逻辑AI工作流相结合导出适当的工作流？
- 除了更高层的网络优化，如何开发基于LLM的物理和无线电链路层，体现LLM的优势？如何使用LLM实现低延迟、超可靠通信？如何对无线通信系统进行优化，以实现LLM在将来的网络上的有效部署与操作？

我们期待着未来有更多的研究工作能在这个领域展开，从而为LLM和网络架构的结合以及内生智能网络的构建带



▲图6 用于意图推理的边缘大型语言模型

² 新必应是指GPT授权的搜索引擎，可通过<https://www.Bing.com/New>访问。

来更多的创新和突破。

参考文献

- [1] OpenAI. GPT-4 technical report [EB/OL]. [2023-08-10]. <http://arxiv.org/abs/2303.08774>
- [2] ZHANG J Y, VAHIDIAN S, KUO M, et al. Towards building the federated GPT: federated instruction tuning [EB/OL]. (2023-05-09) [2023-08-10]. <http://arxiv.org/abs/2305.05644>
- [3] XU M, DU H, DUST N, et al. Unleashing the power of edge-cloud generative AI in mobile networks: a survey of AIGC services [EB/OL].(2023-03-28)[2023-08-10]. <http://arxiv.org/abs/2303.16129>
- [4] WU W, LI M S, QU K G, et al. Split learning over wireless networks: parallel design and resource management [J]. IEEE journal on selected areas in communications, 2023, 41(4): 1051–1066. DOI: 10.1109/JSAC.2023.3242704
- [5] KANDPAL N, ERIC W, COLIN R, et al. Deduplicating training data mitigates privacy risks in language models [EB/OL]. (2022-02-14) [2023-08-10]. <https://arxiv.org/abs/2202.06539>
- [6] LIU P F, YUAN W Z, FU J L, et al. Pre-train, prompt, and predict: a systematic survey of prompting methods in natural language processing [J]. ACM computing surveys, 55(9): 1–35. DOI: 10.1145/3560815
- [7] TOUVRON H, THIBAUT L, GAYTUER L, et al. LLaMA: open and efficient foundation language models [EB/OL]. [2023-08-10]. <http://arxiv.org/abs/2302.13971>
- [8] ZHANG B, SENNRICH R. Root mean square layer normalization [EB/OL]. (2019-10-16)[2023-08-12]. <https://arxiv.org/abs/1910.07467>
- [9] SHAZEE N. GLU variants improve transformer [EB/OL]. (2020-02-12) [2023-08-10]. <https://arxiv.org/abs/2002.05202>
- [10] SU J, LU Y, PAN S, et al. RoFormer: Enhanced transformer with rotary position embedding [EB/OL]. (2022-08-09)[2023-08-12]. <https://arxiv.org/abs/2104.09864>
- [11] SU J, LU Y, PAN S, et al. LoRA: low-rank adaptation of large language models [EB/OL]. (2021-04-20) [2023-08-12]. <https://arxiv.org/abs/2104.09864>
- [12] TAORI R, GULRAJANI I. Stanford alpaca: an instruction-following llama model [EB/OL]. [2023-08-10]. https://github.com/tatsu-lab/stanford_alpaca
- [13] WANG Y, KORDI Y. Self-instruct: aligning language model with self generated instructions [EB/OL]. (2022-12-20)[2023-08-12]. <https://arxiv.org/abs/2212.10560>
- [14] ZHU J H, LI R P, DING G, et al. Aol-based temporal attention graph neural network for popularity prediction and content caching [EB/OL]. (2022-08-18)[2023-08-10]. <https://arxiv.org/abs/2208.08606v1>
- [15] PANG L, YANG C, CHEN D, et al. A survey on intent-driven networks [J]. IEEE access, 2020, 8: 22862 – 22873. DOI: 10.1109/ACCESS.2020.2969208

作者简介



陈宇轩, 浙江大学在读博士研究生; 研究方向为大型语言模型在通信场景中的应用及语义通信。



李荣鹏, 浙江大学信息与电子工程学院副教授、博士生导师; 主要研究方向为智能通信网络、网络智能、网络切片等; 曾入选首批博士后创新人才支持计划, 获得浙江省杰出青年基金项目资助, 并获得吴文俊人工智能优秀青年奖、江苏省科学技术一等奖等。



张宏纲, 浙江大学兼任教授、博士生导师; 长期从事无线通信与网络、人工智能、认知通信、绿色通信、复杂网络等领域的研究; 曾获 2021 年 IEEE 通信学会杰出论文奖、《IEEE Internet of Things Journal》最佳论文奖等; 发表论文 265 余篇, 拥有 IEEE 802.15UWB 国际标准提案 16 项、国际专利 3 项。

6G 新型信息通信网络架构设计



Architecture Design of 6G New Information Communication Network

段晓东/DUAN Xiaodong, 孙滔/SUN Tao, 刘超/LIU Chao,
施南翔/SHI Nanxiang
(中国移动研究院, 中国 北京 100053)
(China Mobile Research Institute, Beijing 100053, China)

DOI: 10.12142/ZTETJ.202305012

网络出版地址: <http://kns.cnki.net/kcms/detail/34.1228.TN.20231016.0927.006.html>

网络出版日期: 2023-10-17

收稿日期: 2023-08-12

摘要: 网络架构是每一代通信网络的核心。5G网络架构的变革为6G服务千行百业奠定了基础。6G网络架构设计需结合5G的经验,支持新场景、新指标、新要素,以简化、高效、灵活为目标进行优化。给出了6G网络架构设计的6个原则和5个维度,并以此为基础进一步提出“三体四层五面”的6G总体架构,以及端到端的全服务化系统架构和灵活按需的分布式自治组网架构。最后讨论了6G与5G网络架构的关系,并给出未来研究方向及相关产业发展的建议。

关键词: 6G; 网络架构; 服务化架构; 分布式自治

Abstract: Network architecture serves as the foundation of communication networks of each generation. The evolution of 5G network architecture is the foundation of thousands of 5G services. The design of 6G network architecture needs to combine the experience of 5G, support the new scenarios, new indicators and new elements, and optimize for the goal of building a more simplified, efficient and flexible network. The six design principles and five dimensions of 6G network architecture are presented. On this basis, a "3 bodies, 4 layers, 5 planes (3-4-5)" 6G overall architecture, an end-to-end holistic service-based system architecture, as well as a flexible and on-demand distributed autonomous networking architecture are presented. The relationship between 6G and 5G network architecture is further explored, and the suggestions for future research directions and the development of related industries are proposed.

Keywords: 6G; network architecture; service based architecture; distributed autonomous network

引用格式: 段晓东, 孙滔, 刘超, 等. 6G 新型信息通信网络架构设计 [J]. 中兴通讯技术, 2023, 29(5): 76–81. DOI: 10.12142/ZTETJ.202305012

Citation: DUAN X D, SUN T, LIU C, et al. Architecture design of 6G new information communication network [J]. ZTE technology journal, 2023, 29(5): 76–81. DOI: 10.12142/ZTETJ.202305012

系统架构是通信网络的骨骼和中枢,是移动通信代际发展的核心标志之一。架构的变革既能带动整体网络的系统性创新,又能牵引全产业链及业务服务的升级换代。网络架构设计的前瞻性、可行性、可靠性、效益性、兼容性,都会直接影响移动通信网络的发展路径和应用成效,必须提前考虑。

1 6G 网络架构设计的背景

6G 网络架构设计,既要考虑新场景新需求的驱动,又要考虑技术发展和网络演进态势。本文从以下 5 个方面阐述 6G 网络架构发展的关键驱动力。

1) 5G 新架构优化需求

以服务化架构 (SBA) 为代表的 5G 架构已经全面走向商用。4 年商用的考验,验证了 5G 网络架构设计的合理性和先进性。

5G 网络架构是革新的一代,实现了基础设施的虚拟化、

云化和网络功能的服务化。新一代移动通信的发展不仅带动了移动通信产业,还增强了运营商网络能力。5G 虚拟化、云化、服务化,改变了运营商传统的运维、运营和管理模式,同时带动了网络的整体转型。5G 商用部署的过程也呈现出若干问题,这些问题都需要在 6G 网络架构设计中进行优化。

(1) 服务化等设计理念实现有待深化。一方面,原子化服务需优化,受到兼容 2G ~ 4G 网络互操作的影响,部分服务尚存耦合性,导致新特性、新服务的引入会影响已有服务;另一方面,服务化机制尚未普及到端到端所有网元,致使部分网元仍被传统接口绑定,不利于敏捷网络目标的实现。

(2) 5G 服务能力与业务深度融合后,展现出更高的需求。例如,随着定位业务的发展,逐步出现行业边缘场景中,园区设备定位及跟踪数据不出园等要求。现有定位方案依赖于集中部署的核心网网元及定位平台,面向随着业务发

展出现的新需求，既需提升数据安全能力，也需满足准实时要求。

(3) 在5G规模商用部署中还发现了用户数据难迁移、服务等级协议(SLA)跨域协作不便利、网络自动化运维能力不高等问题。

2) 6G新场景驱动

新的业务需求是移动通信网络发展的第一驱动力。当前国家电信联盟无线电通信部门(ITU-R)已明确6G六大典型场景^[1]，如图1所示，要求网络内生支持人工智能(AI)、感知、泛在连接等新能力，实现从移动通信向移动信息服务的重大转变，包括沉浸式通信、超大规模连接、极高可靠低时延通信、AI与通信融合、通感一体和泛在连接的六大场景。

3) 6G性能指标要求

性能沉浸化要求网络在支持大规模用户通信和计算的基础上实现灵活弹性组网，并对网络指标提出更极致的要求。业界认为，6G需求指标和5G相比将有进一步提升，控制面时延将达到1 ms，用户面时延将达到0.1~1 ms，连接密度将达到0.01~1亿设备/km²，移动性支持500~1000 km/h，可靠性达到10⁻⁷~10⁻⁵，定位精度将达到1~10 cm^[2-3]。

4) 6G要素融合要求

要素融合化要求网络架构支持通信、感知、算力、智

能、安全五大内生要素全面交叉融合。在传统通信能力上，网络持续增强极致性能体验和全域覆盖，并且在感知的支持下，实现精确感知，推进感知能力变革；在算力的支持下，实现信息和数据的多样泛在处理；在AI大模型的助力下，实现智能化组织管理；在可信和安全内生新体系下，提供更加可靠的安全保障。

5) AI支撑6G网络演进

6G是通信技术与AI技术深度融合的新一代移动通信系统。中国移动在《6G网络架构技术》白皮书^[4]中指出，AI对6G端到端信息处理和服务架构的发展将起到至关重要的作用。通过AI赋能网络，可以降低网络功能原子化、微服务化带来的编排设计复杂性，实现在时变环境下网络服务与业务最优匹配，提升网络运行效率和用户体验。通过网络使能AI，构建端到端的AI学习和推理环境，为AI模型提供可靠传输，提供安全、可保障的AI即服务(AIaaS)，实现AI服务的多样化供给。

2 6G网络架构设计原则

面向6G时代的新场景新需求，6G架构设计应在5G基础上继承并进一步发展，通过多种信息技术深度融合，提供更多要素融合的一体化服务，构建新型信息服务网络。具体来说，6G网络将秉承兼容、跨域、分布、内生、至简、孪生六大设计原则进行架构设计。

1) 兼容设计原则

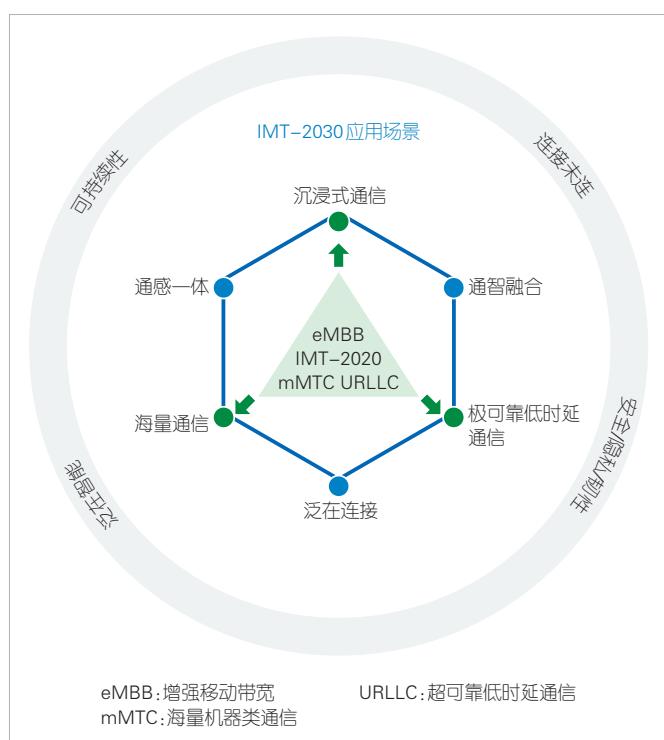
6G网络架构设计将沿着IP化、云化、服务化的方向深化变革，实现前后向兼容。在端到端IP化的基础，进一步实现确定性IP；在云化的基础，进一步实现算网一体；在核心网服务化的基础，进一步实现全域服务化和服务化技术演进。充分延续5G的架构优势，支持5G网络平滑发展演进为6G网络。

2) 跨域设计原则

6G网络架构将面向固定、移动、卫星等多种接入方式，公众/行业、物理/数字等多种网络进行核心网、传输网、接入网跨域跨层一体化设计，在架构层面实现多接入的控制融合、业务融合、管理融合以及网络组织的融合，支持网内不同域的协同，为端到端、全域全网提供质量可保障的信息通信服务。

3) 分布设计原则

6G网络架构将通过集中+分布的协同组网，实现资源、路由、功能等的分布式管理和优化调度，并实现自生长、自优化、自演进的网络自治，从而在大规模复杂组网环境下实现网络资源和网络能力的优化调度，满足6G场景下海量连



▲图1 ITU-R定义的IMT-2030典型应用场景

接和极致性能的要求。

4) 内生设计原则

6G网络架构设计将向内生式转变，在全域全程中内置安全、AI等核心要素，并渗透到各领域、各网络、各单元的全生命周期中，实现上述要素与通信网络最深程度的融合。面向多域融合、连接泛在、资源异构的6G网络，满足大规模组网下的多样化、多元化服务需求。

5) 至简设计原则

6G网络架构设计将支持面向小规模、简化拓扑场景的极简组网，一方面定义基本网络功能集合，整合零散的功能，设计简化网络架构，减少网络复杂度；另一方面构建轻量化服务，细化服务和功能的颗粒，减少服务间的耦合度，并支持智能化的组织能力，降低系统组网及维护难度。

6) 孪生设计原则

6G网络架构设计将向物理+数字孪生方向演进，构建平行的物理和数字网络，形成虚实融合管理手段，一方面支持不同网络和业务形态的实时建模，另一方面支持从数字向物理实体灵活实时的一体化策略控制，基于数字系统实现物理系统的预测和优化^[5]。

3 面向新型信息服务的6G网络架构

3.1 6G网络架构设计的5个维度

基于前文所述的场景和设计原则，6G网络将转变为能够提供通信、感知、计算、智能、大数据、安全等多要素服务的新一代信息服务网络。6G网络架构设计主要包括以下5个维度的考量：

1) 信息全过程服务

6G网络架构将从仅支持传统的信息传输服务，拓展到支持信息感知、传输、存储、处理、利用等全过程服务，实现网络平台化、服务多样化^[6]，使得AI技术贯穿信息流的全生命周期。

2) 服务于数字世界

6G网络架构将从“碳基”转变为“硅基”，依托新型网络基础设施以及算力网络、人工智能、大数据、高速泛在通信、通感互联等体系化的技术能力，实现数字孪生世界。

3) 空天地一体化

卫星互联网成为全球新一轮信息网络建设的焦点。6G网络架构从设计上应当支持天地融合、泛在连接，推动地面通信与卫星通信产业生态深度融合。

4) 确定性网络

6G网络架构需要满足和匹配多样化分级化的业务需求，设计可预测、可度量、可保障的通信底座，满足业务对带宽、时延、抖动、可靠性等的通信性能的确定性需求。

5) 高效能网络

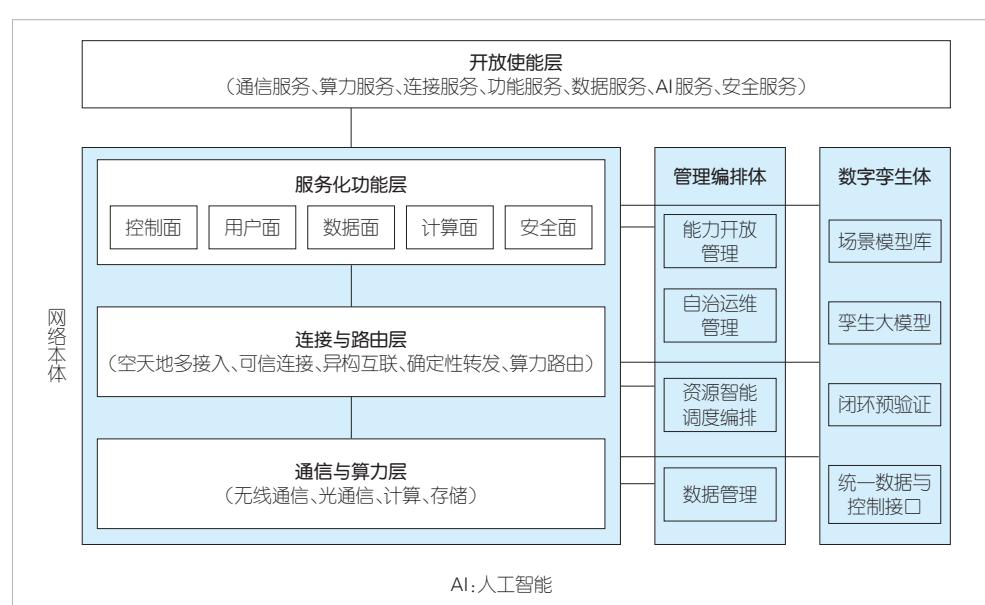
在国家“双碳”战略引导下，6G网络发展需要持续向绿色低碳演进。6G网络架构需要遵循智简设计原则，对信息流和能量流进行系统级协同调度，提升网络整体效能。

3.2 “三体四层五面”6G总体架构

基于上述6G网络架构设计的维度，本文提出“三体四层五面”的6G总体架构设计，从空间、逻辑与功能组成3个角度呈现一个跨域、跨层、多维的6G网络，具体如图2所示。

“三体四层五面”中，“体”是空间视图，描述了网络的物理构成；“层”是逻辑视图，描述了网络的分层架构；“面”是功能视图，沿用传统第3代合作伙伴计划（3GPP）网络中控制面和用户面的“面”概念，描述了网络的功能类别^[7]。

6G网络由网络本体、管理编排体、数字孪生体三大物理实体构成。其中，管理编排体和数字孪生体是面向6G新定义的两个实体。网络本体是最重要的网络实体，实现网络功能和运行；管理编排体对网络进行实例化及变更操作，实



▲图2 “三体四层五面”的6G总体架构

现全生命周期编排管理；数字孪生体构建了网络的数字空间，实现虚实映射。

6G网络逻辑层次自下而上包含通信与算力层、连接与路由层、服务化功能层和开放使能层“四层”。“四层”的设计突出了6G架构在分层要素和能力上的丰富，体现了多域协同、融合发展的理念。通信与算力层的设计突出了“算力”资源要素^[8]，为6G提供频谱、算力、存储、通信融合的基础资源；连接与路由层延续开放协议的设计理念，不断吸收新机制和新协议，向可编程、确定性的方向演进；服务化功能层延续服务化的设计理念，SBA从核心网拓展到全领域，支持不同功能按需灵活构建；开放使能层进一步丰富对外开放的信息和能力，通过提取、封装、编排、组合，为自有业务和第三方应用提供服务。

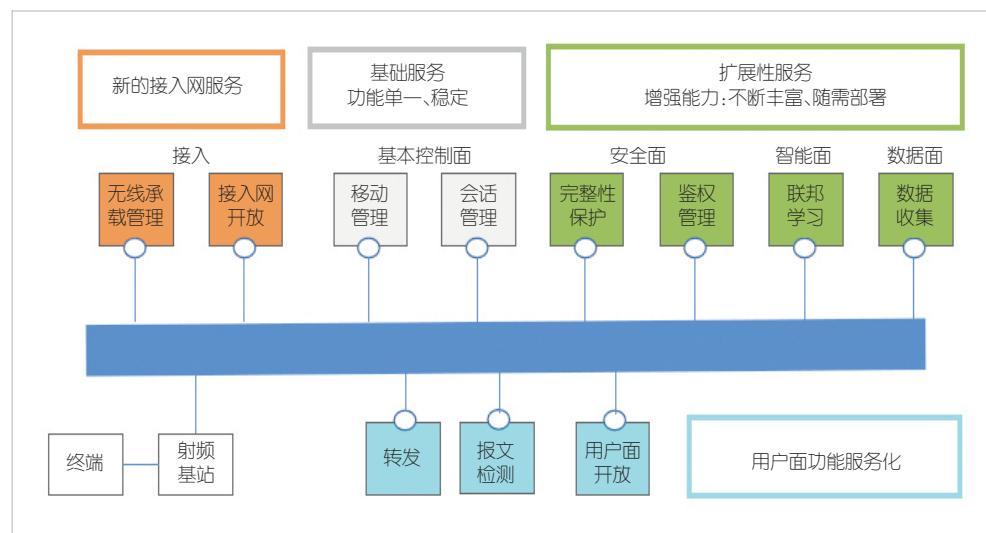
6G网络架构在功能层上大大增强了传统控制面和用户面，并且引入了新的数据面、计算面和安全面，由此共同组成“五面”。其中，控制面向全服务化方向演进，能够实现多种接入方式的融合控制；用户面则向服务化、可编程、确定性方向演进，以实现灵活、高性能的转发功能；新增的数据面，旨在解决用户数据迁移困难的代际难题，系统性地提供可信数据服务；新增的计算面引入了算力要素，能够构建分布式智能节点，协同提供全局AI能力，实现智能内生；新增的安全面，建立以零信任为基线的安全体系，实现安全内生。

3.3 全服务化的系统架构

全服务化架构（HSBA）通过服务定义端到端系统构成^[9]，设计了整个端到端系统的交互形式，包括组件、协议和连接，基于服务化接口进行信息交换和业务处理，体现了“三体四层五面”6G架构的系统设计方法，如图3所示。

全服务化架构采用统一的服务框架技术进行服务自组织，并通过服务化接口进行交互。在全服务化架构中，功能层中的各个面以及每个面内部的各个网络功能都采用模块化、服务化设计。

全服务化架构是服务化架构在多个领域的拓展。终端服务化、接入网服务化和核心网全面服务化是全服务化架构的



▲图3 6G全服务化系统架构(HSBA)

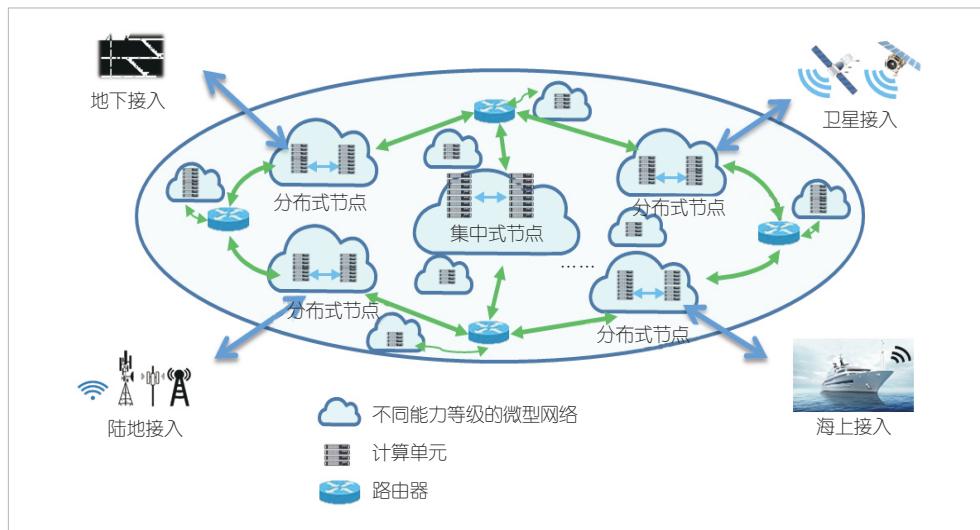
有机组成部分。在服务化功能层，控制面和用户面在5G基础上进一步增强和优化，采用新型服务化机制和调用方法。6G网络架构设计将进一步探索服务解耦机制，通过Service Mesh/Mecha技术，实现业务逻辑和分布式系统之间更加深入的松耦合，提升应用程序编程接口（API）调用效率；引入GraphQL技术，以图的数据结构进行保存，有效提升查询效率。

对于6G新定义的数据面、计算面和安全面，网络架构将采用模块化、服务化设计，实现服务要素拓展。6G网络中引入“算力”和“数据”要素，基于服务化设计理念新增计算面和数据面，为智能服务提供要素基础，真正做到智能化网络与服务化网络的融合。

3.4 分布式自治的组网架构

分布式自治网络架构（DAN）通过灵活、按需、智能的组网设计构成^[10]，呈现了“三体四层五面”6G架构中各个网络单元之间的连接关系和组网形态。DAN由功能自包含、同质化、自闭环的微云单元（SCU）组成，具备完整的系统框架和功能、资源、连接能力，以及自组织、自管理、自优化的能力。6G DAN采用集中加分布的方式部署：分布式部署的SCU遵循至简设计原则，作为靠近用户的前台；集中式部署的SCU功能相对完备，作为靠近业务系统的中后台，具体如图4所示。

SCU是分布式自治网络架构中最关键的模块单元。SCU功能可按照全服务化的方式进行组织，具备本地完成数据和信令处理的能力，能够实现网络高效响应。SCU支持按需定制，其基础设施规格、连接协议、服务化功能、开放能力均



▲图4 6G分布式自治网络架构(DAN)

可以按场景需求进行定制化设计。SCU可以在网络中按需构建，单元之间可快速便捷组网。此外，SCU还具备自治能力，能够实现无人管理、自主运行、自动感知、实时调整，从而满足差异化多样化的业务需求。

微云单元之间的通信采用通用传输协议，支持多SCU之间灵活连接、按需互通，实现SCU即插即用、快速部署。SCU引入统一的控制面和用户面协议，克服2G到5G移动网络中使用多种协议交互带来的复杂度问题。对于用户面协议，可考虑采用与承载网协议融合的用户面协议，例如基于IPv6的段路由（SRv6），不再需要端到端隧道。此时网络可以方便地控制和调整数据包的转发路径。

4 与5G网络架构的关系

1) 6G网络应是5G网络的继承式、迭代式发展。

5G SBA网络架构总体上是创新的，因此在设计6G网络架构时，应将5G SBA网络架构作为6G设计的重要基础。尽管6G是全新一代的网络，支持更多业务场景，提供更多服务能力，但6G的演进是脱胎于5G网络的，是在5G基础上的继承式创新而非颠覆式变革。

2) 6G网络应将实现标准和产业的统一作为设计目标。

在5G阶段，由于各方对网络能力的需求有不同的看法，因此在组网形态方面，最终形成了只提供高速传输的非独立组网（NSA）和能够提供全面新能力的独立组网（SA）两种组网方案。这最终导致了标准和产业的不统一。当前6G起步较早，应当避免5G组网设计中出现的NSA组网方案，尽量将SA作为唯一目标组网方案，实现全球6G架构统一标准。

3) 6G架构设计应考虑融合组网和互操作。

6G网络应支持与5G网络融合组网。一方面，由于6G是在5G基础上的继承式创新，因此6G核心网要在自身架构更具特色的同时，兼容5G核心网的架构和能力。也就是说，6G核心网需要融合5G核心网。另一方面，由于对6G网络能力的进一步需求会引入新的空口技术，例如通感一体、天地一体、太赫兹等，可能会存在与5G基站非融合部署的6G新基站，因此6G架构设计还需要考虑与5G的互操作。

5 总结和展望

体系架构的创新关系到网络总体发展，是6G网络最核心的创新之一。新需求、新场景和新要素的出现，赋予了6G网络架构在传统连接和转发之外的多维能力，也为6G网络架构的创新带来新的驱动力，将显著提升网络能力，为用户提供更加极致和更加丰富的业务体验。因此，本文提出了多维立体的6G总体架构“三体四层五面”，从需求、架构、组网的角度系统阐述中国移动的6G架构总体设计，并提出了对6G架构前瞻的研判、理念和思考。

建议产业界各方进一步关注系统架构对6G技术方向的引领作用，从一开始就大力投入到网络架构的攻关中，并从产业、标准等方面做好准备，与运营商共同完成6G网络架构的设计和迭代更新，加速架构方案的收敛，为全球6G架构统一标准打好基础。

参考文献

- [1] ITU. Framework and overall objectives of the future development of IMT for 2030 and beyond [EB/OL]. [2023-09-10]. <https://www.itu.int/md/R19-WP5D-230612-TD-0905/en>
- [2] ZHANG P, SHEN X M, ZHANG J H. Editorial for the special issue on 6G requirements, vision, and enabling technologies [J]. Engineering, 2022, 8: 1–2. DOI: 10.1016/j.eng.2021.12.001
- [3] UUSITALO M A, RUGELAND P, BOLDI M R, et al. 6G vision, value, use cases and technologies from European 6G flagship project hexa-X [J]. IEEE access, 2021, 9: 160004–160020. DOI: 10.1109/access.2021.3130030
- [4] 中国移动. 6G网络架构技术白皮书 [EB/OL]. [2023-09-10]. <http://221.179.172.81/Images/20220622/36941655883996900.pdf>
- [5] 孙滔, 周铖, 段晓东, 等. 数字孪生网络(DTN): 概念、架构及关键技术 [J]. 自动化学报, 2021, 47(3): 569–582. DOI: 10.16383/j.aas.c210097
- [6] 王晓云, 段晓东, 孙滔. 平台化服务网络: 新型移动通信系统架构研究 [J]. 电信

- 科学, 2023, 39(1): 20–29. DOI: 10.11959/j.issn.1000-0801.2023008
- [7] DUAN X D, WANG X Y, LU L, et al. 6G architecture design: from overall, logical and networking perspective [J]. IEEE communications magazine, 2023, 61(7): 158–164. DOI: 10.1109/mcom.001.2200326
- [8] 中国移动. 算力网络技术白皮书 [EB/OL]. [2023-09-10]. <https://www.digitalelite.cn/h-nd-1936.html>
- [9] WANG X Y, SUN T, DUAN X D, et al. Holistic service-based architecture for space-air-ground integrated network for 5G-advanced and beyond [J]. China communications, 2022, 19(1): 14–28. DOI: 10.23919/jcc.2022.01.002
- [10] WANG S, SUN T, YANG H W, et al. 6G network: towards a distributed and autonomous system [C]//Proceedings of 2020 2nd 6G Wireless Summit (6G SUMMIT). IEEE, 2020: 1–5. DOI: 10.1109/6gsummit49458.2020.9083888

作者简介



段晓东, 中国移动研究院副院长, 正高级工程师, IMT-2030 (6G) 推进组网络技术组组长; 主要研究方向包括5G/6G网络架构、IP新技术、IPv6、算力网络等。



孙滔, 中国移动研究院首席专家, 正高级工程师, 中国科学技术协会第十届全国委员会委员; 长期从事移动通信网架构、IP新技术研究和标准化工作。



刘超, 中国移动研究院主任研究员, 高级工程师; 长期从事TD-LTE、NB-IoT、5G/6G核心网等领域的技术研究和标准化工作。



施南翔, 中国移动研究院主任研究员, 高级工程师; 长期从事5G/6G网络架构、天地一体网络、能力开放、码号等领域的技术研究和标准化工作。



相干光收发器件未来技术演进

Future Technology Evolution of Coherent Transceiver Optical Devices

沈百林/SHEN Bailin,王会涛/WANG Huitao

(中兴光电子技术有限公司,中国南京210012)
(ZTE Photonics Technology Co., Ltd, Nanjing 210012, China)

DOI:10.12142/ZTETJ.202305013

网络出版地址: <http://kns.cnki.net/kcms/detail/34.1228.TN.20231016.0928.008.html>
网络出版日期: 2023-10-17
收稿日期: 2023-08-12

摘要: 相干光收发器件是相干光通信系统的核心器件。相干光通信系统朝着更大容量、更长传输距离、更低成本方向发展,同时相干光收发器件面临一系列新挑战,包括高带宽、多波段、高性能、高可靠性、高集成度、低功耗。相干光收发器件的未来技术演进包括新材料光芯片、先进封装技术、多路集成架构等方面。光芯片将存在多种材料,包括硅光、磷化铟和薄膜铌酸锂,以及基于硅光平台的异质集成技术。光器件将参考采用微电子行业的先进封装技术,以减小芯片间高速电信号的传输距离,降低成本,保证封装可靠性。器件级多路集成可满足未来多波传输架构的需求。

关键词: 相干通信; 光器件; 技术演进

Abstract: Coherent transceiver optical devices are the key devices of coherent optical communication systems. With the development of coherent optical communication systems towards larger capacity, longer transmission distance, and lower cost, coherent transceiver optical devices face a series of new requirements and challenges, including high bandwidth, wide spectral band, high performance, high reliability, high integration, and low power consumption. Future technology evolution of coherent transceiver optical devices includes: new material optical chips, advanced packaging technologies, and multi-channel integrated architectures. A variety of materials will exist in optical chips, including silicon photonics, indium phosphide and thin-film lithium niobate, as well as heterogeneous integration technologies based on silicon photonics platforms. The optical device will refer to the advanced packaging technology of the microelectronics industry to reduce the high-speed electrical signal transmission distance between chips, and ensure package reliability. The device-level multi-lane integration meets the needs of future multi-wavelength transmission architectures.

Keywords: coherent communication; optical components; technology evolution

引用格式: 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,2023,29(5):82-86. DOI: 10.12142/ZTETJ.202305013

Citation: SHEN B L, WANG H T. Future technology evolution of coherent transceiver optical devices [J]. ZTE technology journal, 2023, 29(5): 82-86. DOI: 10.12142/ZTETJ.202305013

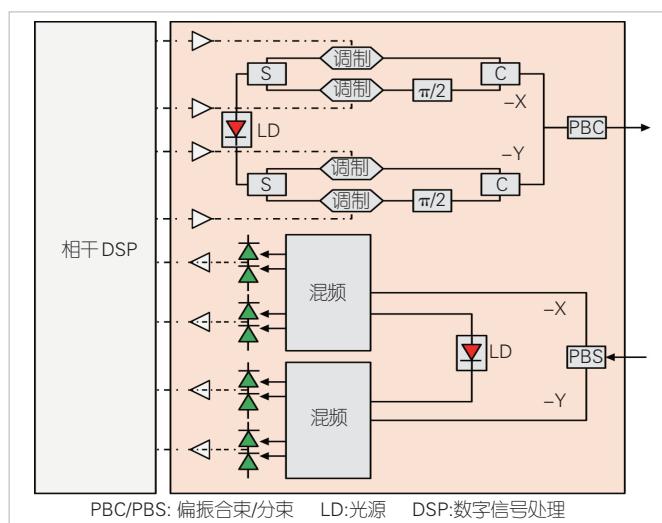
1 相干光收发器件

1.1 具体概念

相干光收发器件^[1-3]是光芯片、电芯片的集成组合。发送部分实现光信号的偏振复用正交调制功能,包含电信号驱动功能;接收部分利用本振信号和输入信号混频干涉,实现双偏振态信号检测和电信号放大功能。相干光收发器件的工作原理如图1所示。

1.2 在系统中的位置

波分复用系统包含光转发单元、复用/解复用单元、光纤、光放单元、可重构分叉复用单元,以及管理单元。光转发单板实现业务信号到波分复用光信号的光电光转换。相干



▲图1 相干光收发器件的工作原理

光模块位于光转发单板的线路侧，其核心是相干光收发器件，如图2所示。

1.3 需求和挑战

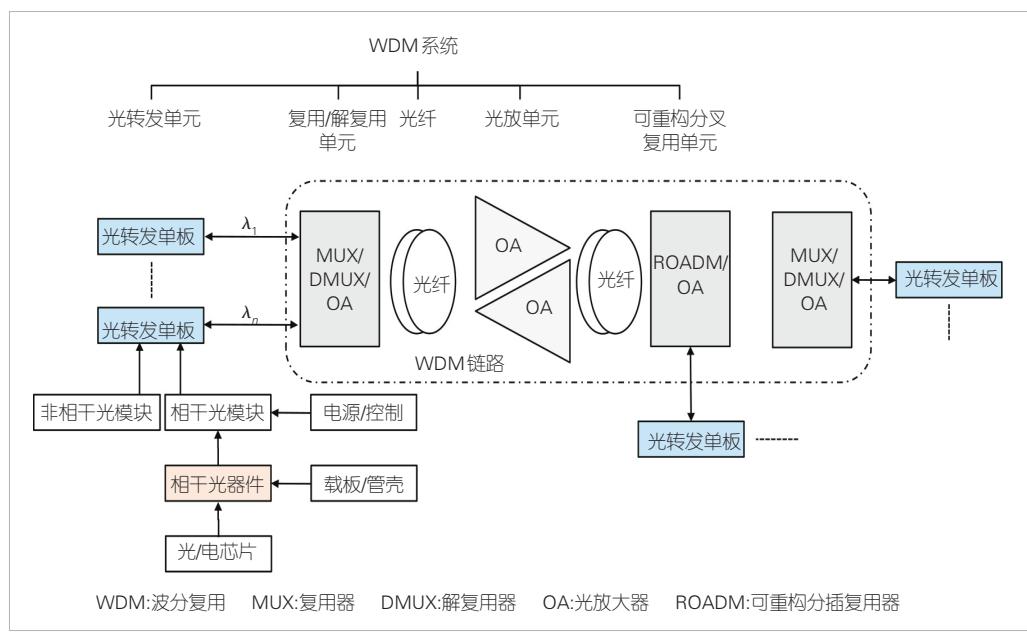
1) 高波特率

自从相干通信技术实现商用以来，信号波特率按照大约3年一代的方式进行演进，如图3所示。数据中心互联的ZR需求（如400ZR/800ZR）使得64、128 GBd相干光收发器件成为主流。128 GBd之后的速率已接近物理极限，技术迭代有可能放缓。业界对下一代信号波特率的观点不统一：主要集中在1.6T-ZR1的波特率选择（192 GBd和256 GBd）方

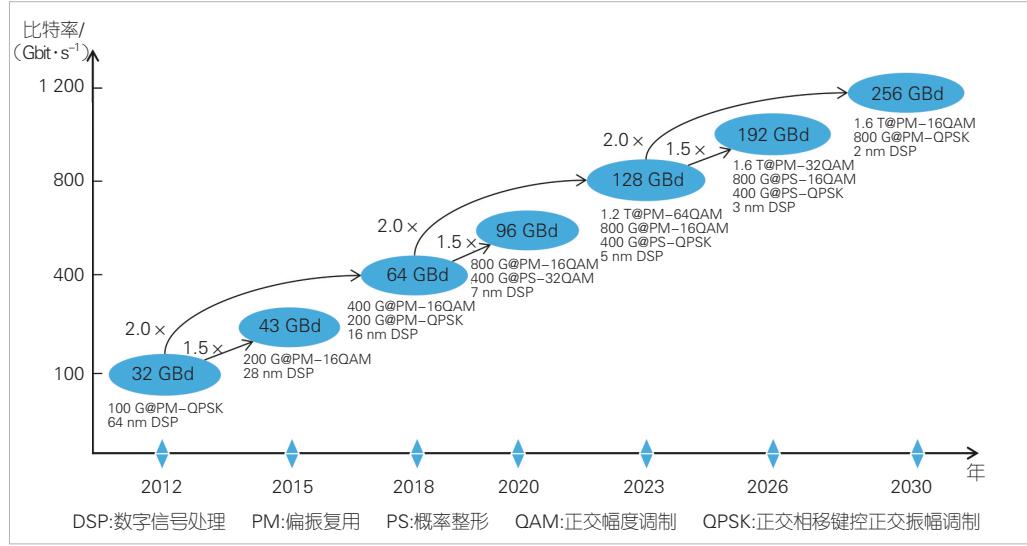
面。其中，256 GBd方案的传输能力强于192 GBd方案，但256 GBd方案存在技术挑战，如数字信号处理（DSP）芯片模数转换（AD）/数模转换（DA）的采样率和带宽，以及模拟电芯片的带宽。同时，256 GBd方案需要配置275 GHz的通道间隔，并且需要扩展波段才能提升系统传输容量；而192 GBd方案只需要配置200 GHz的通道间隔。在G.654.E光纤上，192 GBd方案配合PS-16QAM调制以及更强的前向纠错错误码，可以实现1000 km以上的传输，满足骨干网大部分应用需求，并且系统传输容量有提升。相干DSP芯片的主流晶圆代工厂为台积电。台积电3 nm和2 nm工艺路线支持192 GBd和256 GBd开发。2022年台积电实现3 nm芯片量产，预计2025—2026年量产2 nm芯片。数据中心互联的ZR应用需求远大于长途传输需求。光互连网络论坛（OIF）标准的1.6T-ZR1标准确定的信号波特率影响128 GBd之后的器件带宽选择。

2) 多波段

为保持容量的提升，传统波分复用系统的通道数维持80波。当波特率提升且频谱效率不变时，需要考虑波段扩展。为确保光信噪比（OSNR）性能，骨干网传输系统首选正交相移键控（QPSK）调制方案。100G骨干网波分系统采用C80方案，200G骨干网波分系统采用C120方案，400G骨干网波分系统采用C120+L120方案。800G骨干网波分系统如果仍采用QPSK调制，为实现系统传输容量的升级，需要扩展到S+C+L+U波段；如果采用PS-16QAM调制，需要扩展到S+C+L波段。G.652光纤和G.654.E光纤均支持多波段。S、U波段的相干光收发器件和光放大模块技术成熟度偏低，研



▲图2 相干光收发器件在系统中的位置



▲图3 信号波特率的演进过程

发仍处于实验阶段。多波段系统由于拉曼通道功率转移效应导致 OSNR 不均衡代价，因此需要综合考虑多波段系统的性价比。当波段扩展的代价太大或无法持续扩展波段时，单光纤系统可以不用维持 80 波，此时多芯光纤将成为传输容量升级的另一技术路径。近年来，具有低时延、小非线性效应、低色散、多波段等优点的空芯光纤被广泛认可，但目前距离骨干网规模商用还比较遥远。总之，从技术成熟度考虑，下阶段采用 S+C+L 波段的可能性大于 S+C+L+U 波段，但也不排除考虑性价比继续沿用现有的 C+L 波段。

3) 高性能

相干光模块的 OSNR 性能决定了传输能力，它取决于 DSP 算法、光芯片材料、电芯片性能及光电芯片之间高速电信号互联的质量。骨干网要求传输能力在 1 000 km 以上，属于 OSNR 受限系统，对此一般采用 QPSK 或 PS-16QAM 调制方案。相干光收发器件要求带宽大于所支持最大波特率的一半。从 64 GBd 相干光收发器件开始，电芯片带宽峰化功能有效补偿了光芯片的带宽不足和封装带宽损失。速率越高，器件的实施代价越大。当器件带宽大于 110 GHz 时，封装设计的信号完整性难度大，芯片设计和器件封装将是较大的挑战。

4) 高集成度

小型化、高集成度成为光器件不断追求的基础需求。早期光芯片、驱动器、跨阻放大器、相干 DSP 芯片均为分立形态，现阶段光器件集成了光电芯片以满足小型化需求。2012 年第一代长距相干 100G 光模块采用 300PIN-MSA 模块 (127.0 mm × 177.8 mm)。数据中心短距离相干光模块功耗较低，倾向采用 QSFP-DD 封装 (18.35 mm × 52.40 mm)；电信长途传输应用功耗较大，倾向采用 CFP2 模块 (41.5 mm × 91.5 mm)。光器件尺寸越小，低功耗和先进散热设计就越重要。2012 年第一代长距相干 100G 光模块功耗大于 50 W，2023 年 400ZR 光模块典型功耗不到 18 W，长距 400G 光模块典型功耗约为 25 W，近期研发的 100ZR 的 QSFP28 光模块典型功耗甚至低至 5 W。光模块功耗的降低归因于 DSP 芯片、光源和光器件的技术发展。随着信号波特率提升，DSP 芯片功耗增加，DSP 芯片需要采用更先进的工艺节点来优化尺寸和功耗。

5) 低成本

不断降低成本是相干光收发器件规模应用的需求。波分复用系统升级的驱动力是单位比特传输成本的不断下降。提升信号速率或波特率，减少光电器件数量，有利于降低单位比特传输成本。当技术上信号波特率难以持续提升时，多路集成是可行方案，例如数据中心 1.6 Tbit/s 和 3.2 T bit/s 的短

距 ZR/ER 互联场景。当波段扩展和波特率提升不作为优先需求时，需要考虑多路集成方案。在 400G 时代，相干光模块下沉到 120 km 单跨传输；在 800G 时代，相干光模块下沉到 10 km 和 40 km 单跨传输。相干光模块下沉的前提是非相干技术的传输能力无法满足应用场景需求，同时相干光收发器件的技术发展能够促使成本下降。低成本相干光模块设计思路有：光器件规模化应用和技术发展降低成本、非气密封装代替气密封装、集成驱动器和跨阻放大器的相干 DSP、固定波长光源代替可调谐光源、功能简化版可调谐光源、信号和本振光源同频异纤传输免频率补偿、O 波段传输无需色散补偿、降低采样率降低功耗等。

2 未来技术演进

2.1 新材料光芯片

光芯片材料有硅光、磷化铟和薄膜铌酸锂，以及硅光—薄膜铌酸锂的异质集成新材料，是实现高波特率、多波段、高性能、高集成度、低成本相干光收发器件的关键要素之一。

当前，高速相干光芯片的主流材料有硅光、磷化铟和薄膜铌酸锂。硅光材料支持多波段，具有集成度高、封装友好、成本低等特性，成为可插拔相干模块市场的主流选择。硅锗探测器如果要实现 110 GHz 以上的带宽，就需要减小载流子渡越时间，采用电感带宽峰化措施。一般认为，硅光技术实现 192 GBd 或 256 GBd 仍存在较大的技术挑战。2023 年已有厂商实现相干 128 GBd 硅光相干调制和接收。关于磷化铟器件，早在 2019 年就有研究团队率先公开 192 GBd-QPSK 发射机。近期有多个研究团队实现了 110 GHz 以上带宽的单行载流子 (UTC) 结构磷化铟光电探测器。磷化铟器件存在一些缺点，它需要控制温度，外置偏振管理器件，封装相对复杂，单芯片仅支持单波段工作。传统体材料铌酸锂素有“光学硅”的称号。自 2018 年以来，薄膜铌酸锂材料以大带宽特性成为研究热点。目前已多个研究团队报道了带宽大于 110 GHz 的调制器，预计薄膜铌酸锂调制器将在 128 GBd、192 GBd 和 256 GBd 应用中发挥重要作用。然而，薄膜铌酸锂只能用于发送端，接收端仍需要采用磷化铟或硅光技术。

结合硅光技术和各种高效电光调制新材料的异质集成技术是行业研究热点。这些新材料包括铌酸锂、钛酸钡、有机聚合物等等。硅光技术的缺点是插损大、带宽小，但集成优势也很突出，如果硅光和高效电光调制效应的新材料异质集成，将在未来超 128 GBd 相干光收发器件中发挥重要作用。异质集成技术存在热膨胀系数不匹配、工艺兼容性不好、长

期可靠性不高等问题。现阶段有多个初创公司专注于硅光异质集成钛酸钡或有机聚合物的工艺技术。钛酸钡比铌酸锂材料具有更大的线性电光系数，但钛酸钡的居里温度点很低，制备困难。有机聚合物在高温可靠性、制备工艺兼容性等方面仍需要提升。因此，新材料在商用前需要完成性能、成本、可靠性等多方面的评估和验证。总体而言，当前钛酸钡和有机聚合物技术仍处于实验室阶段，薄膜铌酸锂技术处于工程化阶段。硅光专家预计未来5~10年内硅光商用代工厂的异质集成薄膜铌酸锂工艺将成熟^[4]。硅光-薄膜铌酸锂的异质集成材料优势包括：大带宽(>110 GHz)、多波段(单片支持S+C+L波段)、高性能(铌酸锂OSNR性能好)、高集成度(收发集成)。

2.2 先进封装技术

未来相干光收发器件将采用先进封装技术。这是实现高波特率、高性能、高集成度、低成本相干光收发器件的另一个关键要素。

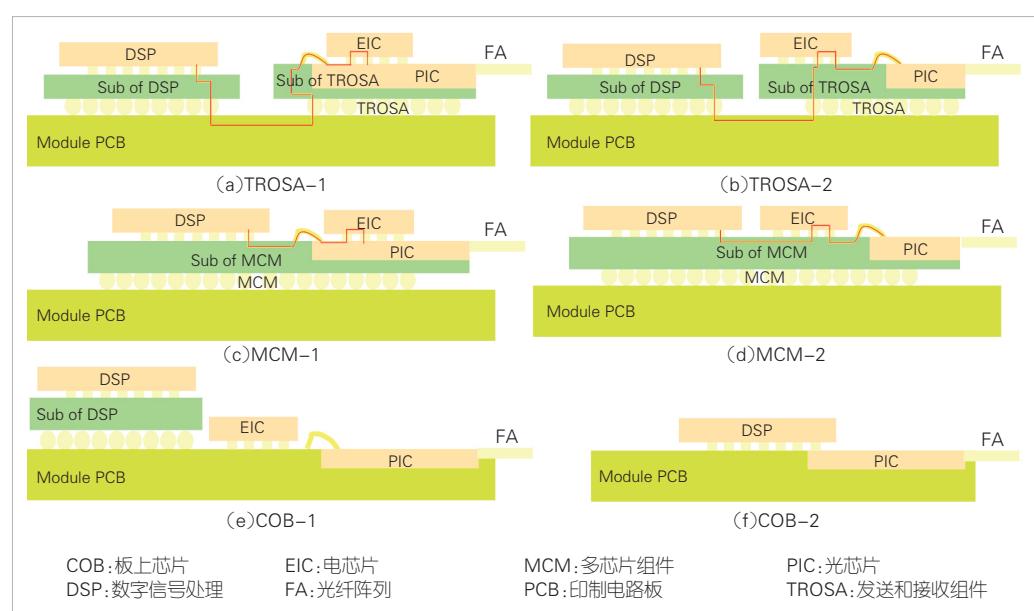
相干光收发器件封装实现光/电芯片集成为一个器件。封装技术^[5]需要综合考虑热、力、电、光的设计，尤其保证器件的长期可靠性和信号完整性。封装方案的设计与可用的光、电芯片接口形态密切相关。目前封装方案主要包括传统的金盒气密封装，以及近年来出现的小型化球栅阵列(BGA)封装和集成DSP芯片的多芯片组件(MCM)封装，未来还可能采用低成本和高性能的板上芯片(COB)封装方案。金盒封装采用高温烧结陶瓷基板。BGA封装多采用有机基板，具有一些优点：尺寸小、加工周期短、成本低、与模块印制电路板(PCB)适配性好、可实现光器件封装微电子化。电芯片散热是设计重点。倒装电芯片要求器件具备从顶部散热的功能。可靠性与基板/管壳材料的选型相关，例如：光器件基板和模块PCB进行BGA焊接时，基板材料的热膨胀系数会影响焊点的可靠性。要优化光器件封装的信号完整性，就需要尽量减少阻抗的不连续点。如图4所示，(c)和(d)的MCM封装架构相对于(a)和(b)的TROSA封装架构，DSP裸片

和光电芯片进行共基板封装，高速电信号少经历了两次由BGA植球引起的阻抗不连续点，改善了高频性能。(e)和(f)均为COB封装。(e)中DSP为封装片，PCB需支持电芯片的倒装，光芯片通过打线与PCB连接。当驱动器摆幅要求不高时，DSP集成驱动器和跨阻放大器成为可能。(f)中DSP为裸片，集成了驱动器(DRV)/跨阻放大器(TIA)功能，且倒装焊跨接在模块PCB和光芯片(PIC)上。需要说明的是，PCB(如类载板)需要高精度工艺支持DSP裸片的倒装。波特率较高时，芯片之间高速电信号的传输路径应尽量短，阻抗不连续点应尽量少，同时DSP芯片应尽量与光芯片共封装。

2.3 多路集成架构

未来相干光收发器件将出现多路集成架构。这是实现高集成度、低成本相干光收发器件的方向之一。

单位比特传输成本影响系统传输方案的选择。在技术发展早期或光电器件带宽难以满足高波特率需求时，多路并行成为技术可行和成本最优的方案选择。相干调制信号的偏振复用正交信号架构相当于非相干调制信号的四路并行或波分架构。在非相干10G、40G、100G、200G、400G、800G、1.6T标准接口中，多路并行架构一直是重要的实现方案。短距离非相干传输采用单纤或少数几根光纤；长距离相干传输采用波分复用技术，其重要指标是传输容量。多路并行方案无法增加系统传输容量，因而早期关注度不高。在相干400G传输的早期阶段，双波方案成为长途传输的可行方案。IEEE 802.3以太网标准正在讨论短距800G/1.6T的相干传输



▲图4 相干光收发器件封装方案示例

技术。有提案建议 1.6T 传输方案采用双波相干 800G。预计 800G 长途传输将采用双波 128 GBd-PDM-QPSK。对于未来 1.6T、3.2T 的长途传输，如果单载波信号波特率的提升不再具备性价比时，双波或四波方案将被采用，从而推动相干光收发器件实现多路集成。

3 结束语

为满足相干光传输系统更大容量、更长传输距离、更低成本的需求，相干光收发器件需要具备更大带宽、更多波段、更高性能、更高集成度、更低功耗等技术特征。未来相干光收发器件技术将向着新材料光芯片、先进封装技术、多路集成架构等方面演进。

参考文献

- [1] OIF. Implementation agreement for IC-TROSA [EB/OL]. (2019-08-20) [2023-08-10]. <https://www.oiforum.com/wp-content/uploads/OIF-IC-TROSA-01.0.pdf>
- [2] 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, 29(4): 78-82. DOI: 10.12142/ZTETJ.202304014
- [3] 王会涛, 张平化, 苏展. 800 Gbit/s 光模块技术及应用 [J]. 中兴通讯技术, 2021,

27(6): 40-46. DOI: 10.12142/ZTETJ.202106008

[4] SUDIP S, BOGAERTS S, CHROSTOWSKI L, et al. Silicon photonics -- roadmapping the next generation [EB/OL]. (2023-05-25) [2023-08-10]. <https://doi.org/10.48550/arXiv.2305.15820>

[5] 刘新阳. 5G 半导体产业发展和创新趋势思考 [J]. 中兴通讯技术, 2021, 27(4): 51-52. DOI: 10.12142/ZTETJ.202104010

作者简介



沈百林, 中兴光电子技术有限公司高级工程师;研究方向为高速光电相干器件技术。



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。

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 (按姓名拼音排序)

| | | | |
|-----|------------|-----|--------------|
| 曹进 | 西安电子科技大学 | 秦志金 | 清华大学 |
| 陈力 | 中国科学技术大学 | 史颖欢 | 南京大学 |
| 陈琪美 | 武汉大学 | 王景璟 | 北京航空航天大学 |
| 陈舒怡 | 哈尔滨工业大学 | 王兴刚 | 华中科技大学 |
| 陈为 | 北京交通大学 | 王勇强 | 天津大学 |
| 官科 | 北京交通大学 | 温森文 | 华南理工大学 |
| 韩凯峰 | 中国信息通信研究院 | 吴泳澎 | 上海交通大学 |
| 何姿 | 南京理工大学 | 夏文超 | 南京邮电大学 |
| 胡杰 | 电子科技大学 | 徐梦炜 | 北京邮电大学 |
| 黄晨 | 紫金山实验室 | 徐天衡 | 中国科学院上海高等研究院 |
| 李昂 | 西安交通大学 | 杨川川 | 北京大学 |
| 刘春森 | 复旦大学 | 尹海帆 | 华中科技大学 |
| 刘凡 | 南方科技大学 | 于季弘 | 北京理工大学 |
| 刘俊宇 | 西安电子科技大学 | 张娇 | 北京邮电大学 |
| 卢丹 | 中兴通讯股份有限公司 | 张宇超 | 北京邮电大学 |
| 陆游游 | 清华大学 | 章嘉懿 | 北京交通大学 |
| 宁兆龙 | 重庆邮电大学 | 赵昱达 | 浙江大学 |
| 祁亮 | 上海交通大学 | 周伊 | 西南交通大学 |
| 秦晓琦 | 北京邮电大学 | 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

中兴通讯技术

(ZHONGXING TONGXUN JISHU)

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（按姓名拼音排序）：

段向阳、高 音、胡留军、华新海、刘新阳、
陆 平、史伟强、屠要峰、王会涛、熊先奎、
赵亚军、赵志勇、朱晓光

双月刊 1995 年创刊 总第 172 期
2023 年 10 月 第 29 卷 第 5 期

总编辑：王喜瑜
主编：蒋贤骏
执行主编：黄新明
编辑部主任：卢丹
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

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

《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：合肥添彩包装有限公司
出版日期：2023 年 10 月 26 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