



中文核心期刊 中国科技核心期刊

第三届国家期刊奖百种重点期刊 信息通信领域产学研合作特色期刊

ISSN 1009-6868

CN 34-1228/TN

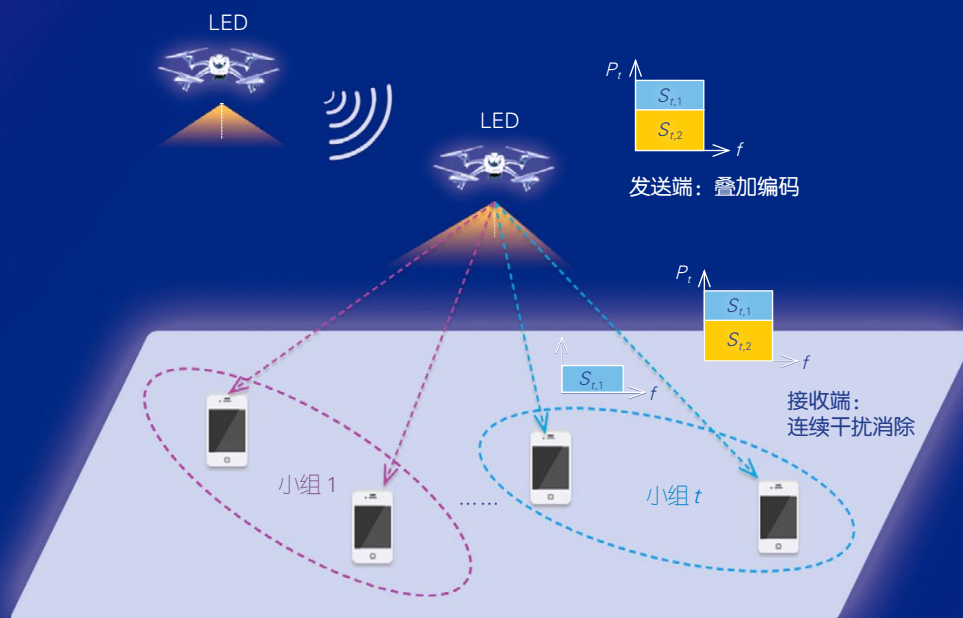
中兴通讯技术

ZTE TECHNOLOGY JOURNAL

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

第 30 卷 · 总第 174 期 · 2024 年 2 月 · 第 1 期

专题：下一代多址接入技术



(封面图片详解见 P15)



《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) 糜正琨(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	陶小峰	北京邮电大学教授
陈前斌	重庆邮电大学教授、副校长	王文博	北京邮电大学教授、副校长
段晓东	中国移动研究院副院长	王文东	北京邮电大学教授
葛建华	西安电子科技大学教授	王喜瑜	中兴通讯股份有限公司执行副总裁
管海兵	上海交通大学教授	王翔	中兴通讯股份有限公司高级副总裁
郭庆	哈尔滨工业大学教授	王耀南	中国工程院院士
洪伟	东南大学教授	王志勤	中国信息通信研究院副院长
黄宇红	中国移动研究院院长	卫国	中国科学技术大学教授
纪越峰	北京邮电大学教授	吴春明	浙江大学教授
江涛	华中科技大学教授	邬贺铨	中国工程院院士
蒋林涛	中国信息通信研究院科技委主任	向际鹰	中兴通讯股份有限公司首席科学家
金石	东南大学首席教授、副校长	肖甫	南京邮电大学教授、副校长
李尔平	浙江大学教授	解冲锋	中国电信研究院教授级高工
李红滨	北京大学教授	徐安士	北京大学教授
李厚强	中国科学技术大学教授	徐子阳	中兴通讯股份有限公司总裁
李建东	西安电子科技大学教授	续合元	中国信息通信研究院副总工
李乐民	中国工程院院士	薛向阳	复旦大学教授
李融林	华南理工大学教授	薛一波	清华大学教授
李自学	中兴通讯股份有限公司董事长	杨义先	北京邮电大学教授
林晓东	中兴通讯股份有限公司副总裁	叶茂	电子科技大学教授
刘健	中兴通讯股份有限公司高级副总裁	易芝玲	中国移动研究院首席科学家
刘建伟	北京航空航天大学教授	张宏科	中国工程院院士
隆克平	北京科技大学教授	张平	中国工程院院士
陆建华	中国科学院院士	张钦宇	哈尔滨工业大学(深圳)教授、副校长
马建国	之江实验室教授	张卫	复旦大学教授
毛军发	中国科学院院士	张云勇	中国联通云南分公司总经理
孟洛明	北京邮电大学教授	赵慧玲	工业和信息化部信息通信科技委常委
石光明	鹏城实验室副主任	郑纬民	中国工程院院士
孙知信	南京邮电大学教授	钟章队	北京交通大学教授
谈振辉	北京交通大学教授	周亮	南京邮电大学教授、副校长
唐宏	中国电信IP领域首席专家	朱近康	中国科学技术大学教授
唐雄燕	中国联通研究院副院长	祝宁华	中国科学院院士

目次

中兴通讯技术 (ZHONGXING TONGXUN JISHU)
第 30 卷 总第 174 期 2024 年 2 月 第 1 期

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

卷首特稿 ▶	01 语义通信:未来通信系统的智简之道..... 张平
热点专题 ▶	下一代多址接入技术
	02 专题导读..... 艾渤, 陈为
	03 6G 新型多址技术探讨..... 严春林, 袁弋非, 王森, 吕思颖
	12 基于无线光通信的非正交多址接入技术研究..... 李亮, 董宇涵, 关迅, 宋健, 张晓平
	19 面向 6G 移动通信的极化多址接入技术..... 崔宏基, 牛凯
	26 无蜂窝大规模 MIMO 中的大规模随机接入..... 胡彦丰, 王东明, 梁楚龙, 尤肖虎
	33 智能增强的免授权多址接入技术..... 于含笑, 费泽松
	41 共享码本随机接入有限块长信息理论极限分析..... 高俊园, 吴泳澎, 张文军
	46 大规模离散 MU-MIMO:低复杂度、信息理论最优检测与多用户编码..... 陈学辉, 池育浩, 刘雷
	55 面向卫星通信与导航的下一代多址接入..... 侯天为, 关达, 孙昕
	60 面向 Critical MTC 的无连接传输..... 李志岗, 袁志锋, 董展谊, 李文斌, 梁楚龙
	68 超大规模天线阵列下的多用户快速波束训练..... 司源, 禹宏康, 陈艺骞
	74 异构大规模分布式网络设计与性能评估..... 叶新泉, 卢光延, 陈艺骞
专家论坛 ▶	82 下一代多址接入技术的挑战与关键进展..... 陈为, 艾渤
企业视界 ▶	89 数据中心光模块技术及演进..... 张平化, 王会涛, 付志明

《中兴通讯技术》2024 年热点专题名称及策划人

1. 下一代多址接入技术

上海交通大学教授 艾渤
上海交通大学教授 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 金石
上海交通大学教授 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 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
中国联通研究院首席科学家 唐雄燕

MAIN CONTENTS

ZTE TECHNOLOGY JOURNAL
Vol. 30 No. 1 Feb. 2024

Guest Paper ►	01 Semantic Communication: the Intelligent and Concise Way to the Future Communication System ZHANG Ping
Special Topic ►	New Generation of Multiple Access Technology
	02 Editorial AI Bo, CHEN Wei
	03 Next Generation Multiple Access for 6G YAN Chunlin, YUAN Yifei, WANG Sen, LYU Siying
	12 Non-Orthogonal Multiple Access Technology Based on Optical Wireless Communication LI Liang, DONG Yuhan, GUAN Xun, SONG Jian, ZHANG Xiaoping
	19 Polar Coded Multiple Access Technology for 6G Mobile Communication CUI Hongji, NIU Kai
	26 Massive Random Access in Cell-Free Massive MIMO System HU Yanfeng, WANG Dongming, LIANG Chulong, YOU Xiaohu
	33 AI-Enhanced Grant-Free Multiple Access Technologies YU Hanxiao, FEI Zesong
	41 Fundamental Limits of Massive Random Access with A Common Codebook and Finite Blocklength GAO Junyuan, WU Yongpeng, ZHANG Wenjun
	46 Massive Discrete MU-MIMO: Low-Complexity Information-Theoretically Optimal Detection and Multi-User Coding CHEN Xuehui, CHI Yuhao, LIU Lei
	55 Next Generation Multiple Access for Satellite Communication and Navigation Networks HOU Tianwei, GUAN Da, SUN Xin
	60 Connection-Free Transmission for Critical MTC LI Zhigang, YUAN Zhifeng, DONG Zhanyi, LI Wenbin, LIANG Chulong
	68 Fast Beam Training for Multiuser on Extremely Large Aperture Array SI Yuan, YU Hongkang, CHEN Yijian
	74 Heterogeneous Large-Scale Distributed Network Design and Performance Evaluation YE Xinquan, LU Guangyan, CHEN Yijian
Expert Forum ►	82 Challenges and Key Progress in Next Generation Multiple Access Technology CHEN Wei, AI Bo
Enterprise View ►	89 Data Center Optical Module Technology and Its Evolution ZHANG Pinghua, WANG Huitao, FU Zhiming

期刊基本参数: CN 34-1228/TN*1995*b*16*98*zh*P*¥20.00*6500*15*2024-02

敬告读者	本刊享有所有发表文章的版权, 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。
------	---

语义通信: 未来通信系统的智简之道



◎ 张平/中国工程院院士、本刊编委



随着5G商业化的加速,6G逐渐成为全球各国着力部署的方向。面向2030年及未来,人类社会将进入智能化时代。社会服务均衡化、高端化,社会治理科学化、精准化,社会发展绿色化、节能化将成为未来社会的发展趋势。从移动互联、万物互联,到万物智联,6G将实现从服务于人、人与物,到支撑智能体高效联接的跃迁。通过人-机-物-智慧的全面互联、协同共生,6G将满足经济社会高质量发展需求,服务智慧化生产与生活,助力人类走进虚拟与现实深度融合的全新时代,最终实现“万物智联、数字孪生”的美好愿景。

遵循移动通信产业“用一代、做一代、看一代”的发展规律,6G的研究率先在学术界展开。全球通信技术发达的国家和地区,如美国、欧盟、中国、日本和韩国等都出台了相应的6G研发政策,并推出了一系列举措。

近年来,在中国政府的大力推动和支持下,在国家中长期发展的战略需求驱动下,为保持并扩大中国5G的领先优势,中国正在积极探索6G自主创新的技术路线。

6G网络有望将人和拥有不同智能程度的机器相互连接起来,使得人-机-物-智慧4类通信对象之间产生大量不同形态的数据。由于各种对象之间的通信不再仅是比特数据的传输,语义和有效性成为“万物智联”不可忽视的重要因素。与语法通信不同,语义通信泛指在不同的智能体间进行以“达意”为目标的通信,其核心在于将语义在通信双方之间的准确传递设为目标。与传统语法通信不同,语义通信并不要求信息的大量符号级准确传递,而是关注发送端输入的语义信息与接收端恢复出的语义信息之间的匹配,通过减少信息交流和理解的时间,提升通信节点协作的效能。这项技术被学者认为有望突破基于经典信息论的通信系统传输瓶颈,解决基于数据的移动通信系统中存在的人机不兼容、难

互通等问题,最终实现“万物智联”。

在语义通信基础理论和关键技术方面,目前我们已做了大量的原创性研究,并取得了令人鼓舞的成果。其中,针对语义通信的全新架构,我们提出模型驱动的智简通信系统;通过设计新型智简发射机与接收机,实现模型构建、信源信道编码一体化能力及信源信息的高效重建、恢复。针对利用终端的智慧能力,将模型作为智简信息传输的核心方式,我们提出了面向智简通信系统的云-边-端协同模型部署与传输方案;基于模型分割、剪枝、蒸馏等技术,提出了智简模型的切分技术,通过安全高效的模型传输机制,为6G终端提供“定制化”智简模型,实现了智简模型灵活增强与高效能传输,并在无人系统、智能电网、视频应用和高轨卫星等场景开展了试验验证,试验结果得到学术界和产业界的共同认可。

我们创新性地提出了模分多址技术(MDMA)。MDMA可解释为将语义特征作为用户专用的地址码,即用语义特征区分用户。传统多址技术考虑时、频、功、空等物理资源来实现多用户区分,而模分多址是从更高信息维度上利用了信源语义域特征实现多用户的区分。理论分析与仿真结果均证实了MDMA技术的诸多优势,也受到了全球学术界及产业界的极大关注。MDMA可以促进通信与智能的全面融合,突破传统代际更迭,推动通信系统的范式跃迁和历史性变革。

通过在5G时代形成的行业、产业、生态联盟等优势,6G仍需凝聚产学研用各方优势,形成合力,从而强化6G关键技术赋能传统产业转型升级。

1G—3G移动通信技术由欧美发达国家提出并长期主导,特别是在多址、编码等核心技术层面。这使得中国一直处于技术围堵和利益压榨的困境。目前,中国6G核心技术攻关和标准研发制定等面临新的机遇和挑战。5G发展经验与教训将为6G演进奠定坚实基础,未来6G将利用语义带来信息利用的跃迁、将智简作为未来通信和网络发展的准则、将MDMA作为未来6G网络的智简新空口技术来驱动移动通信系统的范式演进,从而实现人-机-物-智慧互联及更深层次

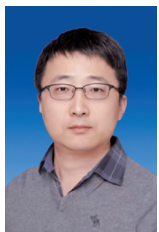
下一代多址接入技术专题导读



专题策划人



艾渤，北京交通大学教授、电子信息工程学院院长，国家6G技术研发总体专家组专家，国家自然科学基金委创新群体带头人，国家杰出青年科学基金、优秀青年科学基金、牛顿高级学者基金、中国科协“求是杰出青年奖”获得者、中共中央组织部“万人计划”领军人才，北京市优秀教师，IEEE Fellow，IEEE VTS杰出讲师；入选斯坦福2020—2022年全球前2%顶尖科学家终身影响力榜，获省部级科技奖励9项。



陈为，北京交通大学教授；长期从事无线通信、信号处理、人工智能技术研究，并围绕高维信息感知和处理、海量机器类通信、语义通信、智慧交通开展研究工作；作为负责人承担国家优秀青年科学基金等多项省部级及以上项目；获2022年度中国信息通信领域十大科技进展、首届詹天佑铁道科学技术奖秦驰道专项奖（青年奖）等学术奖励。

随着无线网络连接设备数量的显著增长，以及万物互联、虚拟/增强现实等多样化网络业务的蓬勃发展，实现超大规模连接、超高带宽效率、超低时延通信迫在眉睫。为满足上述需求，比传统正交多址技术更具优势的新型多址技术成为下一代移动通信技术的研究热点。本期“热点专题”和“专家论坛”栏目共收录了12篇文章，针对下一代多址技术发展所面临的关键挑战和核心问题展开讨论，并介绍下一代多址相关新技术及其与大规模天线、免授权随机接入、无线光通信等结合的相关研究成果。

《6G新型多址技术探讨》回顾了5G非正交多址技术的标准化历程、教训及特点，介绍了6G新型多址的最新理论成果；《基于无线光通信的非正交多址接入技术研究》介绍了无线光通信的非正交多址技的基本原理、系统结构，及其与无人机、智能反射面等新兴技术融合的相关研究进展，并使用强化学习优化了一种新型无人机辅助无线光通信非正交多址方案；《面向6G移动通信的极化多址接入技术》介绍了有源、无源极化多址接入系统，其中包括码域、功率域有源极化多址接入方案，以及一种极化时隙ALOHA无源极化随机接入方案；《无蜂窝大规模MIMO中的大规模随机接入》针对以用户为中心的无蜂窝大规模多输入多输出架构下的随机接入方案展开研究，提出了一种大规模随机接入方案；《智能增强的免授权多址接入技术》介绍了免授权多址接入技术的发展及研究现状，分别分析了结合人工智能技术优化码本与接收机设计、活跃用户检测等方面的免授权非正交多址接入方案；《共享码本随机接入有限块长信息理论极限分析》

针对超大规模机器通信场景中的共享码本随机接入问题展开信息理论研究，在给定约束条件下推导并仿真验证了平均每用户误差概率的非渐近可达界和逆定理；《大规模离散MU-MIMO：低复杂度、信息理论最优检测与多用户编码》研究了实际通信系统约束下的大规模离散多用户多输入多输出系统，提出了低复杂度、信息理论最优接收机和多用户码设计准则；《面向卫星通信与导航的下一代多址接入》面向卫星通信与导航系统，分别对基于多功能、多资源和多技术的下一代多址接入技术展开研究，分析了现有下一代多址接入技术对卫星通信与导航系统的性能增益，总结了结合下一代多址接入技术的卫星通信与导航系统面临的挑战；《面向Critical MTC的无连接传输》提出了一种面向高可靠低时延机器通信的无连接传输方案，并进行了仿真分析；《超大规模天线阵列下的多用户快速波束训练》研究了基于哈希函数的快速、低复杂度新型波束训练方案，该方案具有高准确性；《异构大规模分布式网络设计与性能评估》提出了一种面向6G的异构大规模分布式网络，并介绍分层的分布式网络结构与多链路协同传输方案。“专家论坛”栏目文章《下一代多址接入技术的挑战与关键进展》从海量随机接入、非正交多址两方面介绍了下一代多址技术的关键挑战及进展，分析了两者分别与人工智能、大规模多输入多输出等技术结合面临的挑战与研究成果。

本期的作者来自知名高校与科研机构，针对下一代多址技术，从技术挑战、标准化进展、与各种关键技术结合等方面介绍了最新的研究成果。希望本期内容能为读者提供有益的参考和启示，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！

DOI: 10.12142/ZTETJ.202401002

收稿日期: 2024-01-17

6G 新型多址技术探讨



Next Generation Multiple Access for 6G

严春林/YAN Chunlin¹, 袁弋非/YUAN Yifei²,
王森/WANG Sen², 吕思颖/LYU Siying²

(1. 中关村泛联移动通信技术创新应用研究院, 中国 北京 100080;
2. 中国移动通信有限公司研究院, 中国 北京 100053)

(1. ZGC Institute of Ubiquitous-X Innovation and Applications, Beijing
100080, China;
2. China Mobile Communication Research Institute, Beijing 100053, China)

DOI: 10.12142/ZTETJ.202401003

网络出版地址: <http://kns.cnki.net/kcms/detail/34.1228.tn.20240223.1925.004.html>

网络出版日期: 2024-02-26

收稿日期: 2024-01-10

摘要: 6G 大规模通信场景迫切需要使用新型多址技术解决海量终端随机接入问题, 其他多种场景也需要 6G 新型多址来提高传输速率, 降低接入时延, 增强传输可靠性。回顾了 5G 非正交多址的标准化历程, 总结了各种非正交多址技术的特点, 以及在标准化过程中的经验, 研究并评估了稀疏交织多址和基于 Reed-Muller 码的无源多址两种新型多址性能。本研究可为未来 6G 新型多址的选型和标准化提供有益参考。

关键词: 非正交多址接入; 6G 新型多址接入; 压缩感知; 稀疏交织多址接入

Abstract: The 6G massive machine type communication urgently requires new multiple access technology to support random access of massive machine type terminals, and other scenarios in 6G systems also need new multiple access to increase the data rate, reduce transmission delay and enhance transmission reliability. The standardization process of 5G non-orthogonal multiple access and the properties of various non-orthogonal multiple access techniques are reviewed. The experience in the standardization process is then introduced. The performance of two new types of multiple access is studied, including sparse interleave division multiple access and unsourced multiple access based on Reed-Muller code. This study can provide useful references for the scheme selection and the standardization of the 6G multiple access technology in the future.

Keywords: non-orthogonal multiple access; 6G new multiple access; compressed sensing; sparse interleaver multiple access

引用格式: 严春林, 袁弋非, 王森, 等. 6G 新型多址技术探讨 [J]. 中兴通讯技术, 2024, 30(1): 3-11. DOI: 10.12142/ZTETJ.202401003

Citation: YAN C L, YUAN Y F, WANG S, et al. Next generation multiple access for 6G [J]. ZTE technology journal, 2024, 30(1): 3-11. DOI: 10.12142/ZTETJ.202401003

非正交多址接入 (NOMA) 是无线空口重要基础技术之一, 它一改前几代移动通信系统以正交方式为主的多址方式, 使得同一小区的多个用户可以在相同的时频域和空域资源上叠加传输, 有望大幅提升系统容量, 解决碰撞冲突问题, 因此一直受到学术界和工业界的广泛关注^[1-3]。5G 时代第 3 代合作伙伴计划 (3GPP) 启动了 NOMA 的研究项目, 深入研究了 NOMA 在增强移动宽带 (eMBB)、大规模机器通信 (mMTC)、超高可靠超低时延通信 (URLLC) 三大场景的潜在能力^[1-2]。

2023 年 6 月, 国际电信联盟 (ITU) 完成了国际移动通信 (IMT) 面向 2030 及未来发展的框架和总体目标建议书, 提出 6G 的六大使用场景, 对 5G 中的 eMBB、mMTC、URLLC 场景进行演进^[4]。为满足这些场景的需求, 需要研究新一代的多址技术。

在 ITU 的目标建议书发布之前, 中国学术界和工业界对

6G 应用场景进行了分析和预测^[5-7], 其中包括海量连接场景、密集紧要连接场景、空天地一体化场景和大容量场景, 如图 1 所示。海量连接场景主要针对物联网, 这是因为物联网终端数量巨大, 每平方公里就有千万个终端需要连接。密集紧要连接场景主要针对工业互联网, 它对通信时延和可靠性有很高的要求。空天地一体化场景需要考虑卫星和地面终端接入, 服务终端数量大, 信号往返时延大, 卫星移动速度快。大容量场景主要是针对增强现实 (AR)、虚拟现实 (VR)、扩展现实 (XR) 等极高速数据传输场景。为进一步提高连接密度, 降低接入时延, 提高传输可靠性, 为移动用户提供空天服务, 我们需要研究新型多址技术。

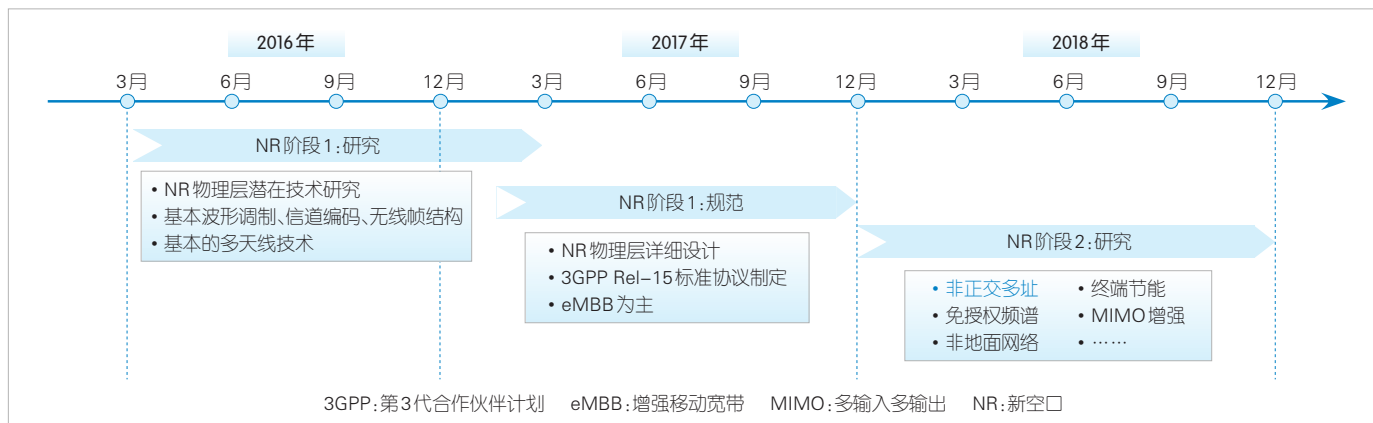
表 1 给出了海量连接场景、密集紧要连接场景、空天地一体化场景和大容量场景在 5G 时代的能力, 以及 6G 时代使用新型多址需要解决的问题。在以上多种场景中, 海量连接场景最具挑战性且需求紧迫。新型多址在海量连接下的研究



▲图1 新型多址接入的四大应用场景^[5]

▼表1 新型多址接入技术需求指标^[5]

应用场景	5G 的能力	设计目标示例	面临的挑战
海量连接	一个宏基站支持每 10 ms 内完成 300 个终端的多址接入和小数据包传输	一个宏基站支持每 10 ms 内完成 3 000 个终端的多址接入和小数据包传输	需要支持超多用户随机接入、低碰撞概率导频设计、用户激活检测和信道估计,需要海量用户数据传输和检测方法
密集紧要连接	目前商用 5G 最大可支持 400 个左右的无线资源控制(RRC)连接;5G 的通信时延是 1 ms;不能同时满足 5G 三大场景需求	在单个基站覆盖范围内,同时满足下述指标:同时服务 2 000 个以上的设备、平均数据流量达到 20 Gbit/s 以上、通信时延小于 0.1 ms、丢包率小于 10 ⁻⁵ 等	对发射信号进行比特交织和比特重复(或扩频)处理,需要新的混合自动重传设计、信号功率激增设计等
空天地一体化	单载扇可以支持 400 个左右的 RRC 连接;控制面时延可能将占到总时域资源的 20% 以上,明显降低了系统的效率	每波束需要同时服务 5 万个窄带物联网设备,明显降低控制面的总时延,把控制面和数据面的空口交互次数降至 2~4 次,把信令交互的资源开销降至 60%~80%	需要低时延随机接入方案,存在多用户干扰时大多普勒频率偏移估计和补偿等问题
大容量	要达到 1 Tbit/s 的峰值数据速率和 5 Gbit/s 的用户体验数据速率是非常困难的	达到 1 Tbit/s 的峰值数据速率、5 Gbit/s 的用户体验数据速率	需要非正交多址传输方案、高速率信道编码调制器,以支持多用户快速收敛迭代检测



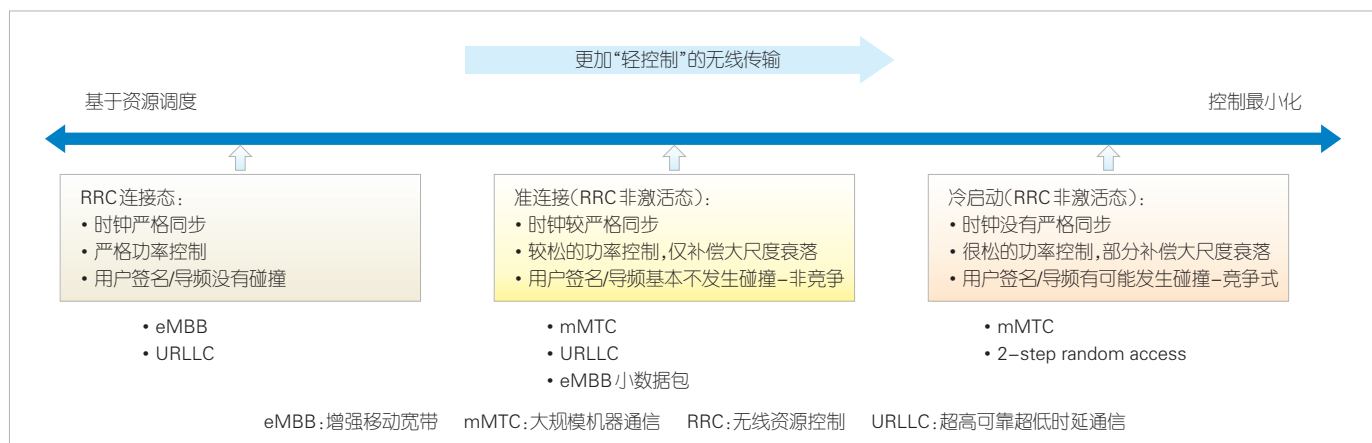
▲图2 3GPP NR研究历程

成果也可以应用到其他场景,因此,海量连接下新型多址将是本文关注的重点。

1 5G 非正交多址研究

5G 系统支持三大主要应用场景,即 eMBB、mMTC 和 URLLC。NOMA 可以实现较高的频谱效率、较多的用户接入数量、较低时延以及高可靠传输。在 5G 时代中国提出了多用户共享接入(MUSA)、图样分割多址接入(PDMA)、稀疏码分多址接入(SCMA)和交织多址接入(IDMA)等技术方案^[8-11]。其中,MUSA 使用复数扩频序列来区分用户,PDMA 使用不同的 pattern 来区分用户,SCMA 采用不同的星座图映射及资源映射来区分用户,IDMA 使用交织器来区分用户。3GPP 在 5G 时代对 NOMA 进行了研究,研究历程如图 2 所示。

无限资源控制(RRC)连接态是 NOMA 研究中的一个重要特性。不同类别的 RRC 连接态如图 3 所示。RRC 连接态有严格同步和功率控制两种,同时用户签名和导频是由系统调度的,可以避免碰撞。某些 eMBB 和 URLLC 场景属于这种分



▲图3 三大场景与RRC状态关系

类。当用户处于RRC非激活态（准连接）时，时钟严格同步，功率控制较松，用户签名和导频基本不碰撞。mMTC、URLLC和eMBB小包传输属于这种类型。在RRC空闲态（冷启动），时钟没有严格同步，只有开环功率控制，此时用户签名或导频会发生碰撞。mMTC和2-step random access属于这种类型。NOMA研究的主要场景是RRC非激活态和RRC空闲态下的mMTC、URLLC和eMBB小数据包业务。在NOMA研究过程中，为避免与NB-IoT的部署场景重叠，单用户的频谱效率不能低于0.1 bit/Hz。这种限制给NOMA研究带来了不利影响，限制了mMTC用户同时接入数。

5G新空口（NR）物理层的基本处理模块，包括波形调制、信道编码等已在NR阶段1完成。这使得NR非正交多址的设计空间大大受限。NOMA的波形调制采用的是正交频分复用（OFDM）或离散傅里叶变换扩频正交频分复用（DFT-S-OFDM），信道编码是低密度奇偶校验码（LDPC）。mMTC、URLLC和eMBB这三大场景在小包传输时采用的码长较小，如50~200 bit。LDPC的优势主要体现在码长较大的情形，如大于1 000 bit，而小码长时采用LDPC会有较大的性能损失。

NOMA传输方案包括以下几种典型方案。符号线性扩展类的传输方案用线性扩展序列区分用户。该方案直接在调制符号上扩展，具有较高的可扩展性和灵活性。扰码/交织不用于区分用户。沿用经典调制方式的有正交相移键控（QPSK）等，eMBB场景的信道编码方式有5G NR LDPC码等。典型接

收机为最小均方误差-连续干扰消除（MMSE-SIC）。该接收机较依赖最小均方误差（MMSE）解调器来抑制用户间的干扰。此外，接收机也可以使用基于MMSE的迭代接收机^[12]。该传输方案支持采用面向竞争式免调度的设计^[13]。

扰码/交织类的传输方案扰码/交织用于区分用户，具有较高的可扩展性和灵活性。该方案沿用经典调制方式和eMBB的信道编码方式。典型接收机为：匹配滤波器/最小均方误差（MMSE）的基本信号估计（ESE）+软比特输出的译码，较为依赖迭代译码解调。扰码/交织类传输方案的缺陷是缺乏面向竞争式免调度的设计。

调制与扩展联合优化类的传输方案符号调制与符号扩展联合设计，直接从编码比特映射到扩展的调制符号。该方案沿用传统信道编码方式和5G NR LDPC码。典型接收机为期望传播算法（EPA）+软比特输出的译码，较为依赖迭代译码解调，其缺陷是缺乏竞争式免调度的设计。

5G NOMA多址方案的分类如表2所示。

5G NOMA研究最终没有形成标准化，原因在于：

- 没有明确的场景需求，尽管有每平方公里百万连接数的指标，但同时接入的用户数不是很多，并未真正体现海量接入的挑战性。
- 缺乏基础理论的突破，尤其在短数据包下的竞争式免调度传输。
- 所研究的免调度传输主要是非竞争式的，研究相对透彻，但相关设计不适合竞争式免调度。

▼表2 非正交多址方案及其特点

方案类别	多址签名数量	竞争式免调度机理	邻区干扰抑制能力
线性扩展类	32~196	无导频的盲检测	有一定的抑制能力
扰码/交织类	没有限制	缺乏相关设计	困难
调制扩展联合设计类	12~24	缺乏相关设计	困难

- 对比方案是低码率方案 LCRS 且接收天线较多时, 在并发用户数不高的情况下 NOMA 方案相比 LCRS 方案增益有限。

- 对系统级仿真中的邻区干扰认识不足, 很多方案存在性能瓶颈。

随着无线通信需求的发展, 6G 时代对新型多址的研究热潮再次掀起。与 5G 时代每平方公里百万连接不同, 6G 时代连接密度是 5G 的 10~100 倍^[5-7]。使用传统的正交多址难以实现如此高的连接密度, 5G 时代研究的 NOMA 也难以满足要求, 因此 6G 时代需要研究新的多址技术。最近通信理论的突破也为 6G 多址提供了很好的理论基础。5G 时代复用的用户数一般不超过 15, 当前所研究的 6G 新型多址支持的用户数可达 300, 远大于 5G 时代的指标。同时 6G 研究考虑了竞争式免调度, 其导频设计及信号检测都与 5G 时代有很大的不同。因此, 5G NOMA 未能完成的使命将在 6G 时代重新启动。

2 6G 潜在新型多址技术

6G 多址与当前学术界的一些最新研究成果密不可分, 因此, 这里我们将介绍当前的最新理论成果。

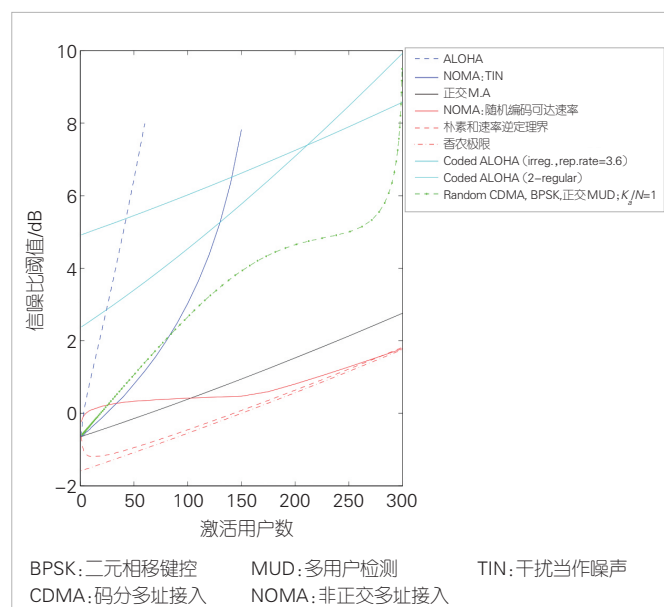
2.1 多址理论的最新成果

假设一个高斯多址接入信道有 K_a 个用户处于激活状态, 即正在进行上行传输。这些用户到基站的信道增益 (路径损耗和小尺度信道衰落) 均为 1, 噪声为高斯白噪声。与传统的多用户信息论不同, 这里所有用户均采用同一个编码码本。同时差错事件是对每个用户定义的, 而不是全体用户的联合差错事件。另外, 有限码长的因素也需考虑。这些定义比较适合免调度情形下的海量小包业务传输。

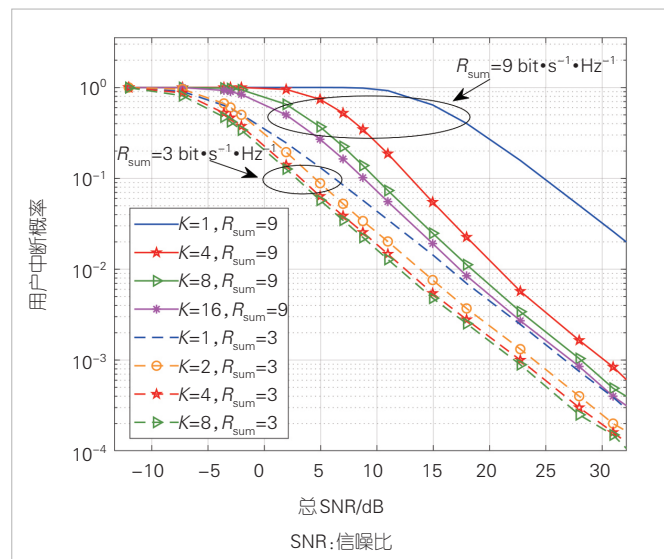
图 4 比较了不同传输方案的理论信噪比阈值, 以及与激活用户数的关系^[14]。在经典 ALOHA 类型方案中, 由于存在碰撞, 信噪比阈值随激活用户数的增加而急剧增加。非正交单用户方案接收机需事先知道激活用户使用的签名, 不是严格意义上的无物理层标识的传输。将其他用户信号的干扰当作噪声 (TIN) 的非正交单用户方案在用户数小于 120 时的性能优于经典 ALOHA。匹配滤波之后再行干扰消除 (多用户检测) 的非正交单用户方案, 在用户数较大时的性能也优于经典 ALOHA 方案, 但信噪比阈值较大。正交单用户传输性能虽然较好, 但这需要事先划分好资源, 不适合随机接入和数据传输的场景。非正交随机多用户编码是严格意义上的无物理层标识的传输。当激活用户数 K_a 低于 100 时, 非正交随机多用户编码稍微逊于正交单用户传输; 但当用户数超

过 100 之后, 非正交随机多用户编码的性能优于正交单用户传输, 并且这个差距会随着用户数的增加进一步扩大。

图 5 是 Rayleigh 衰落多址信道在总的频谱效率一定的情况下, 不同复用用户数下的平均中断概率^[15]。其中 K 是用户数, $K=1$ 的情况相当于正交传输, 只不过系统把所有的资源都分给了一个用户。随着用户数的增多, 每个用户的频谱效率下降, 用户间的干扰也愈明显, 但是总的信噪比要求在逐步降低。从平均中断概率的分析结果可以看出, 相比于 $K=1$ 的单用户情形, 无物理层标识的非正交传输通过降低每个用户频谱效率提高了多用户迭代检测的性能, 从而提高了多用



▲图 4 高斯多址接入信道不同传输方案的理论信噪比阈值与激活用户数的关系^[14]



▲图 5 Rayleigh 衰落多址信道的用户平均中断概率^[15]

户系统的频谱效率。多用户的增益随着系统负载（即用户数或者总频谱效率）的增大而变得更加显著。

压缩感知（CS）是学术界近年提出的信号处理方法，可有效检测稀疏信号的数值^[16-20]。如果信号是稀疏的，可由远低于采样定理要求的采样点重建恢复。压缩感知的核心是信号在某个变换域是稀疏的，可以用一个与变换基不相关的观测矩阵将变换所得信号投影到另外一个信号空间上，通过不断的迭代检测来完成对原始信号的精确估计。压缩感知理论为新型多址导频设计、检测提供了理论基础，可用于多用户激活检测及信道估计。

2017年，Y. POLYANSKIY给出了无源多址接入（UMA）传输方案，支持300个用户的随机接入，与5G时代的十几个并发用户相比，大大提高了接入用户数^[21]。同时，此方案包含了导频设计、用户激活检测、信道估计，以及多用户数据检测。此方案引入了业界最新的压缩感知技术，用于用户激活检测和信道估计。通过使用大的压缩感知矩阵，支持的码本可高达 2^{15} 。在 2^{15} 码本随机选取300用户时，用户碰撞概率很低。此方案一经提出，就引发了极大关注。后来又有学者提出稀疏IDMA传输方案，以实现300用户的随机接入^[22]。稀疏IDMA方案利用压缩感知矩阵进行导频设计。将 $2^{15} \times 2^{15}$ 的FFT矩阵行随机交织并截短成 2000×2^{15} 矩阵，可以实现300用户的激活检测和信道估计。通过在数据传输部分使用IDMA进行用户区分，同时在数据里加入大量零来有效降低用户间的干扰，稀疏IDMA可以较好地实现大规模用户接入和数据传输。下面我们介绍两种潜在6G新型多址方案并给出初步性能评估。

2.2 6G新型多址方案

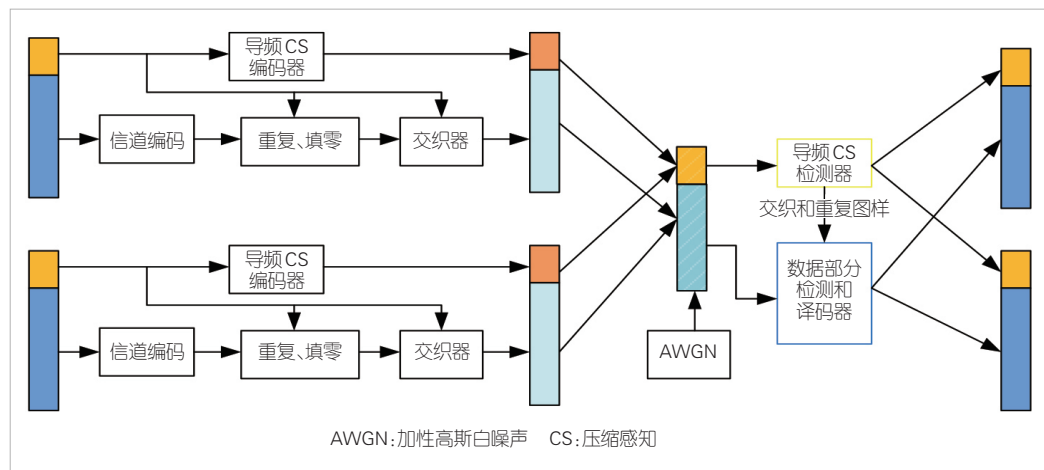
2.2.1 单时隙传输方案

稀疏IDMA+压缩感知是海量用户无源随机接入的重要技术方案^[22]。稀疏IDMA方案的核心设计思想是结合两项技术。第一项技术是基于压缩感知的导频编码。为了支持无物理层标识传输，该技术需要指示用户独特信息，可以将交织图样、比特重复次数、填零数目等信息组成一个很大的码本。通过码

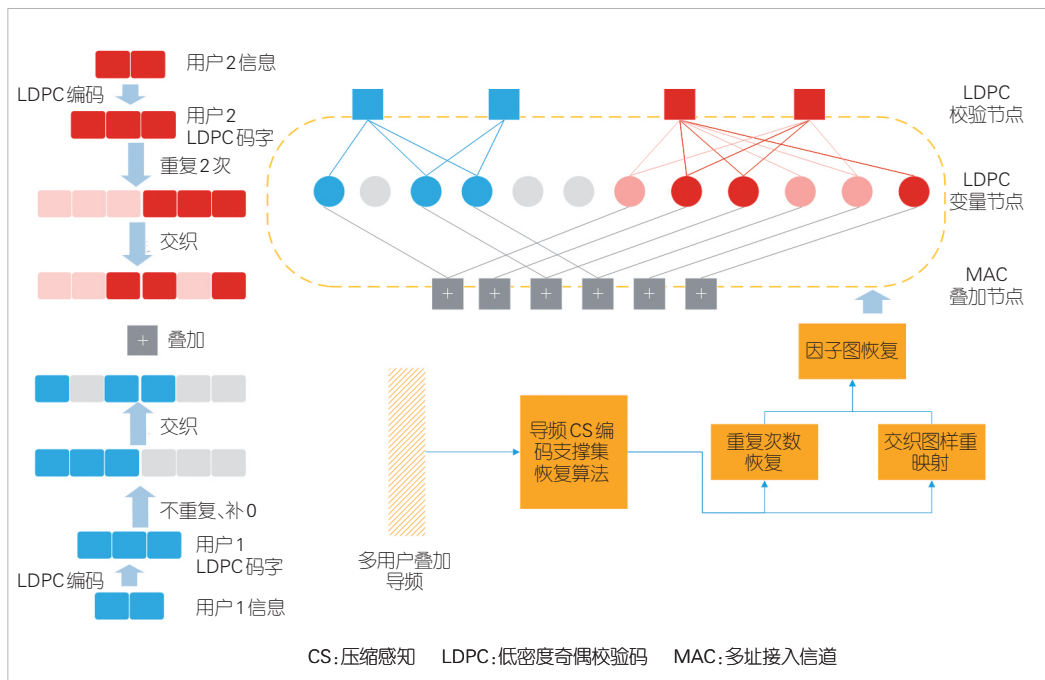
本中码字序号区分不同用户的独特信息。码本序号经过压缩感知映射成为较短的导频，附加在数据部分之前，组成复合包进行传输。第二项技术是稀疏IDMA叠加编码。该技术采用重复、填零、交织、叠加方式，通过比特重复来提高抗多用户干扰的能力，加入大量零元素来有效降低多用户干扰，使用不同交织器来区分用户并随机化多用户干扰。使用这些技术可以大大提高多用户检测的性能。

图6展示了稀疏IDMA的系统架构，该架构包括两个部分：导频编码和稀疏IDMA编码。导频编码的一种方式是采用快速傅里叶变换（FFT）矩阵，将FFT正交矩阵的行随机交织后打孔，得到一个维度较小的矩阵作为导频编码集合。在接收端，导频与IDMA码字分开译码，先通过导频恢复交织图样、比特重复次数、填零数目，然后在已知的叠加因子图结构上进行置信传播（BP）或ESE译码，最终将对应的两部分信息译码结果拼合得到用户的完整发送信息。以两个用户多址接入信道（MAC）系统为例，稀疏IDMA的编译码原理如图7所示。用户1和用户2都采用同样的LDPC码进行编码（这里以LDPC编码为例），它们各自的校验节点到变量节点的因子图是相同的。用户1经过LDPC编码之后没有重复，只是补0，所以因子图中相应部分的边数没有增加。用户2经过LDPC编码之后重复2次，因子图相应部分的边数加倍。比特交织之后，因子图的边分布进一步随机化。两个用户分别的因子图通过MAC叠加节点联系起来，构成一个三层的整体因子图。整个因子图的配置信息，包括重复次数和交织图样，都通过导频的压缩感知恢复算法解出，以辅助IDMA的BP译码。

稀疏IDMA信道编码也可以采用其他编码方式，如卷积码、不规则重复累积码和多元低密度奇偶校验码（NBLDPC）。由于使用了迭代检测，信道编码的译码器需要



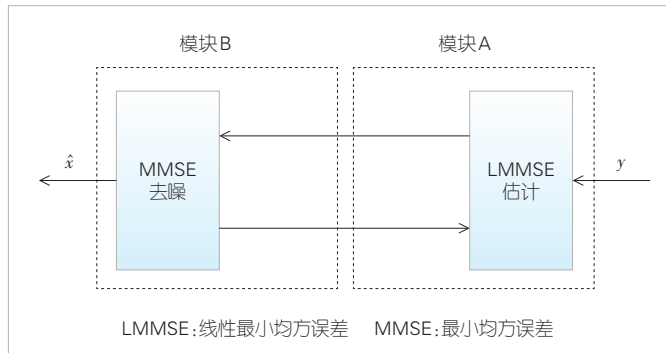
▲图6 稀疏交织多址接入(IDMA)的系统架构^[22]



▲图7 稀疏交织多址接入(IDMA)的编解码原理

提供软入软出的译码信息。极化码难以提供软入软出的译码信息，或提供的软入软出译码信息有较大的性能损失，一般在稀疏IDMA里不使用。使用密度演进可以计算出用户使用不同重复次数的性能（度分布优化），并找出最优的重复次数。

图8是一个用Turbo压缩感知检测的例子。Turbo压缩感知包含两部分。模块A是一个线性MMSE估计器，输入是接收信号 y 和模块B的输入。模块B通过合并 x 的先验信息和模块A的输入进行MMSE去噪。信息在两个模块间不断迭代以得到更精确的 x 值。迭代次数与用户数、导频长度和感知矩阵维度的大小有关。控制迭代次数可以控制计算的复杂度。降低感知矩阵的维度也可以有效降低计算复杂度。AMP算法存在某些条件下不收敛的缺点，基于此正交近似消息传递（OAMP）被提出。通过正交化输入估计误差和输出估计



▲图8 基于压缩感知的迭代检测

误差，OAMP在奇异信道（如相关信道等）取得比AMP更优的性能。但OAMP里有MMSE矩阵求逆操作，复杂度较大。通过在记忆近似消息传递（MAMP）中引入记忆，严格的正交性使得估计器错误成为渐进独立同分布的高斯信号。在MAMP中当前输出估计错误与所有输入估计错误正交。MAMP不需要矩阵求逆也可以取得和OAMP一样的性能，因此具有较好的应用前景。

2.2.2 多时隙传输方案

Reed-Muller码可以生

成高维度的正交矩阵。如同高维度的FFT矩阵，Reed-Muller码生成的正交矩阵也可以用于压缩感知^[23]。这种正交矩阵在无源多址中可以用于用户激活检测和信道估计。利用Reed-Muller码的无源多址方案如图9所示。其中， K 表示用户信息比特， $K-2p$ 个比特用于Reed-Muller码的编码。

基于Reed-Muller码的无源多址方案使用 2^p 时隙传输多用户数据。用户 $2p$ 个信息比特可用于确定传输时隙的序号（总的时隙数是 2^p ）。为提高用户检测成功概率，每个用户在两个时隙上传输。两个时隙同时都检测失败的概率较低，因此，这种策略可以提高用户检测成功的概率。当一个用户信息在时隙A被成功检测，其在时隙B的信号将被干扰消除，以增加其他用户在时隙B的检测成功概率。我们可以使用用户信息比特的最后 $2p$ 个比特确定用户在两个时隙传输的序号，此时用户用于Reed-Muller码编码的比特数是 $K-2p$ 。Reed-Muller码的用户激活检测和数据检测的算法描述如算法1所示^[24]。

算法1：码元重构算法

Input: measurements $y \in \mathbb{R}^n$, where $n = 2^m$; target sparsity S .

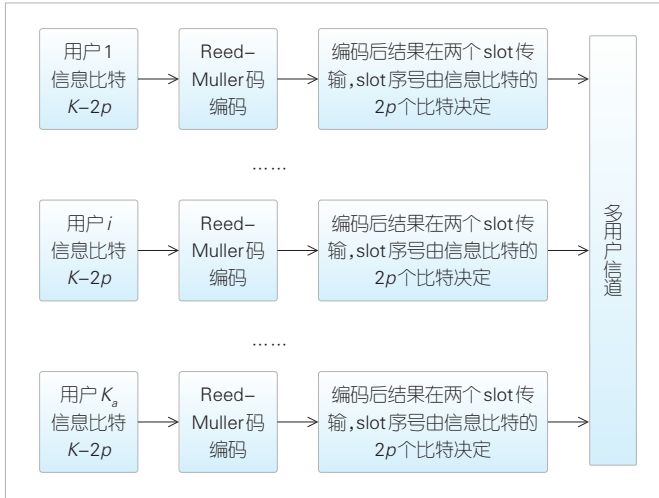
Initialization: $\Phi = []$; iteration count $s = 0$.

While $\|y\| > \varepsilon$ and $s \leq S$ do

$s \leftarrow s + 1$

$(P_s, b_s) \leftarrow \text{findPb}(y)$

$\Phi \leftarrow [\Phi \ \phi_{P_s, b_s}]$



▲图9 基于Reed-Muller码无源多址方案^[24]

$$c_s \leftarrow \Phi^{-1} y$$

$$y \leftarrow y - \Phi c_s$$

end while

Output: $\{P_s, b_s, c_s\}$

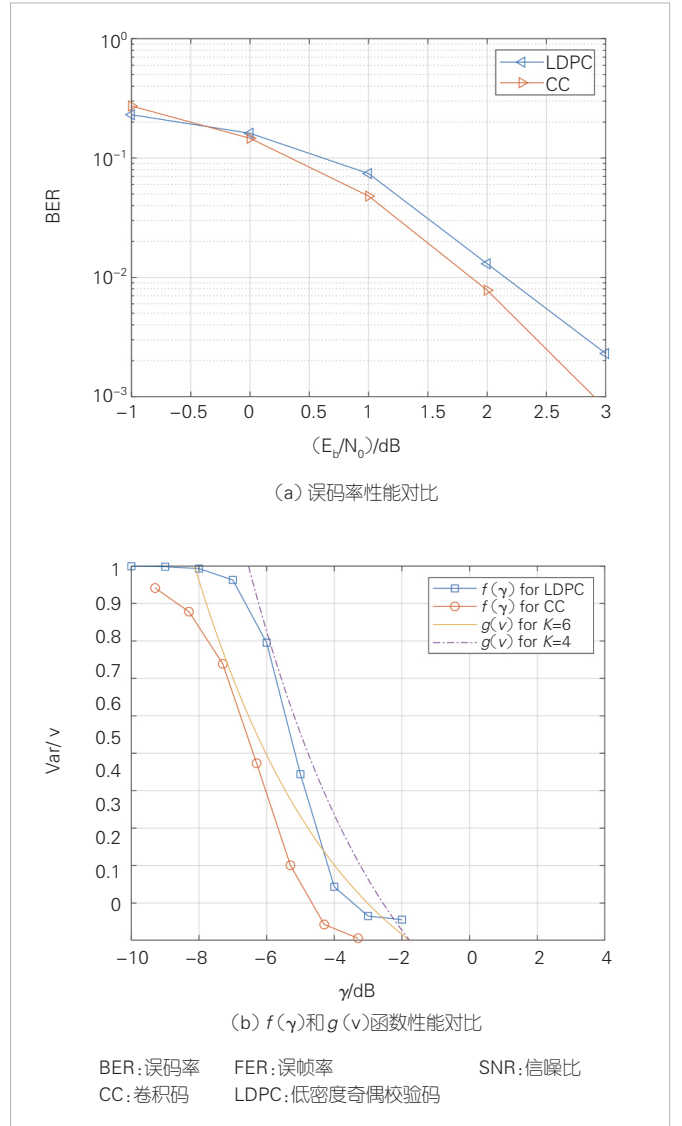
其中, y 是接收信号, m 是Reed-Muller码的阶数, 编码后的数据长度是 2^m 。findPb(y)是Reed-Muller码根据接收信号 y 的Reed-Muller码译码, 译码后得到矩阵 P_s 和矢量 b_s 。 $\phi_{p,b}$ 是根据 P_s 和 b_s 进行Reed-Muller编码后的结果。 Φ 是长度为 2^m 的Reed-Muller码组成的矩阵, C_s 是信道增益, Φ^{-1} 是对矩阵 Φ 求伪逆。

3 仿真结果

表3给出了仿真参数设置, 如信道、信道编码、码率、调制、信息比特长度、编码后长度以及激活用户数。图10(a)给出了单用户使用LDPC码和卷积码的性能。可以看出, 卷积码在误码率(BER)小于0.1后有明显的性能增益。

▼表3 仿真参数配置

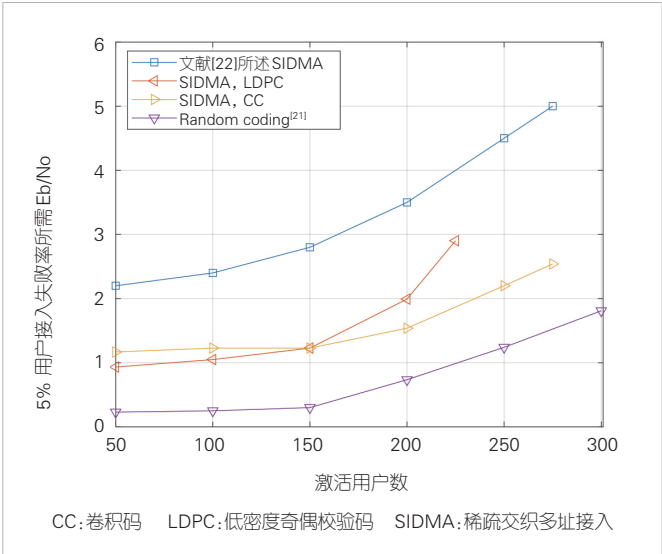
参数	配置
信道	加性高斯白噪声
调制	二元相移键控
信道编码	5G低密度奇偶校验码、卷积码
卷积码生成多项式	[133, 171] _{8进制}
重复次数	2
信道编码码率	0.5
压缩感知导频长度	2 000
多用户数据传输长度	28 000
压缩感知信息比特长度	13
每用户数据部分信息比特长度	87
激活用户数	50 ~ 275



▲图10 卷积码和LDPC性能

卷积码的BER更小意味着用于多址接入时多用户干扰更小, 因此可以取得更好的性能。图10(b)给出 $f(\gamma)$ 函数和 $g(v)$ 函数。稀疏IDMA在每个用户传输的信号里填入了大量的零, 大大降低了用户间干扰。我们认为IDMA系统的用户数 $K=6$ 对应稀疏IDMA系统的激活用户数是250, IDMA系统的用户数 $K=4$ 对应稀疏IDMA系统的激活用户数是200。当 $K=6$ 时, 卷积码的 $f(\gamma)$ 函数不与 $K=6$ 的 $g(v)$ 函数相交, 表明此时使用卷积码可以迭代收敛; 而 $K=6$ 时, LDPC码的 $f(\gamma)$ 函数与 $K=6$ 的 $g(v)$ 函数在 v 接近1的时候相交, 这表明此时使用LDPC不能迭代收敛。当 $K=4$ 时, LDPC码的 $f(\gamma)$ 函数与 $g(v)$ 函数在 v 接近0.05的时候相交, 表明此时使用LDPC可以迭代收敛。

图11给出了达到 $P_e=0.05$ 所需要的 E_b/N_0 。与文献[22]中性能相比, 由于我们使用了FFT更小的压缩感知矩阵, 我们



▲图11 门限随活跃用户数的变化关系

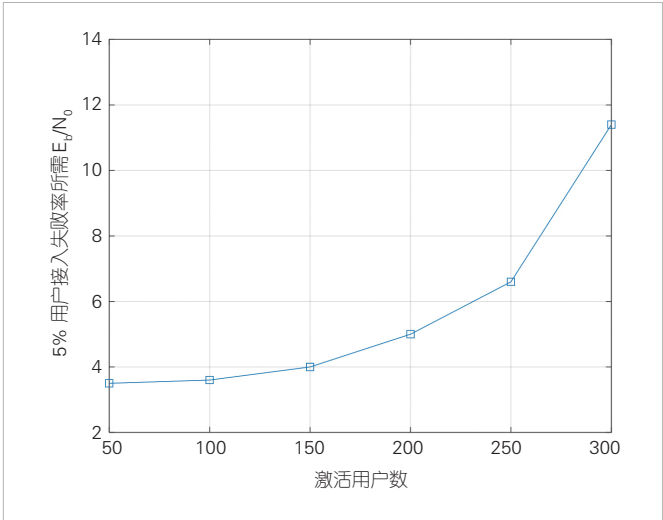
仿真的稀疏交织多址（SIDMA）性能更好（同样是LDPC码）。由于我们的SIDMA没有对重复因子进行优化，取值都是2，SIDMA使用LDPC时在用户较大时有明显的性能损失。同样是SIDMA方案，使用CC码性能比LDPC更好。图中在Eb/No小于3 dB时使用CC可以支持275个用户，使用LDPC仅仅支持225个用户。这说明信道编码对SIDMA性能有较大的影响。我们仿真结果（使用CC的SIDMA）与理论界有大概1 dB的性能损失。此处仿真所有用户重复因子是2。在进行重复因子的优化后，SIDMA可以取得更好的性能，如支持的激活用户数更大，或需要的信噪比更低。

表4给出了基于Reed-Muller无码源多址方案性能仿真的参数设置。仿真结果如图12所示。可以发现，当时隙数是128时，Reed-Muller无源多址方案可支持300个用户。激活用户数大于200后，由于用户间干扰加剧，5%用户接入失败率所需比特信噪比增加较快，但在激活用户数小于200时基于Reed-Muller码无源多址方案性能较好。Reed-Muller码的检测仅仅涉及相关运算，计算复杂度较低，因此适合作为信息比特数较小（42 bit左右）的低复杂度无源多址方案。

▼表4 基于Reed-Muller码的无源多址仿真参数

信道	AWGN
调制	BPSK
时隙数	128
激活用户数	50~300
信息比特数	42
编码后比特数	256
编码方法	Reed-Muller

AWGN:加性高斯白噪声 BPSK:二元相移键控



▲图12 基于Reed-Muller码无源多址方案性能

4 结论

本文中我们回顾了5G NOMA的研究历程，总结了NOMA主要方案及其特点，以及5G NOMA标准化的经验，同时介绍了当前大规模通信的理论突破和6G新型多址的特点。6G新型多址采用竞争式免调度设计，支持海量用户接入，导频设计和信号检测使用了新的范式。特别地，我们介绍了两种典型方案：稀疏IDMA和基于Reed-Muller码的多址方案。这两种方案都可以支持接近300用户的随机接入。针对稀疏IDMA，我们讨论了信道编码对其性能影响。在稀疏IDMA中使用卷积码可以取得比LDPC更好的性能。信息比特数较少时，Reed-Muller码可以满足大规模用户随机接入，并且计算复杂度较低。综合来说，稀疏IDMA有更大的设计灵活性，支持不同的用户数、不同长度的信息比特和不同类型的信道编码，具有更好的应用前景。

参考文献

[1] 3GPP. Study on non-orthogonal multiple access (NOMA) for NR: 38.812 [S]. 2018

[2] YUAN Y F, YUAN Z F, TIAN L. 5G non-orthogonal multiple access study in 3GPP [J]. IEEE communications magazine, 2020, 58(7): 90-96. DOI: 10.1109/MCOM.001.1900450

[3] YAN C L, YUAN Z F, LI W M, et al. Non-orthogonal multiple access schemes for 5G [J]. ZTE communications, 2016, 14(4): 11-16. DOI:10.3969/j.issn.1673-5188.2016.04.002

[4] ITU-R. Framework and overall objectives of the future development of IMT for 2030 and beyond [R]. 2023

[5] 中国移动通信集团有限公司. 6G新型多址接入技术白皮书 [R]. 2023

[6] IMT-2030(6G)推进组. 6G典型场景和关键能力白皮书 [R]. 2022

[7] IMT-2030(6G)推进组. 6G总体愿景与潜在关键技术白皮书 [R]. 2021

[8] YUAN Z F, YU G H, LI W M, et al. Multi-user shared access for

- Internet of Things [C]//Proceedings of IEEE 83rd Vehicular Technology Conference (VTC Spring). IEEE, 2016: 1–5. DOI: 10.1109/VTCSpring.2016.7504361
- [9] DAI X M, CHEN S Z, SUN S H, et al. Successive interference cancelation amenable multiple access (SAMA) for future wireless communications [C]//2014 IEEE International Conference on Communication Systems, IEEE, 2014: 222–226
- [10] NIKOPOUR H, BALIGH H. Sparse code multiple access [C]//2013 IEEE 24th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC), 2013, IEEE: 332–336
- [11] PING L, LIU L H, LEUNG W K. A simple approach to near-optimal multiuser detection: interleave-division multiple-access [C]//Proceedings of IEEE Wireless Communications and Networking, 2003. IEEE, 2003: 391–396. DOI: 10.1109/WCNC.2003.1200381
- [12] YAN C L, YUAN Y F. Spreading based multi-branch non-orthogonal multiple access transmission scheme for 5G [C]//Proceedings of IEEE 89th Vehicular Technology Conference (VTC2019-Spring). IEEE, 2019: 1–5. DOI: 10.1109/VTCSpring.2019.8746411
- [13] YUAN Z F, YAN C L, YUAN Y F, et al. Blind multiple user detection for grant-free MUSA without reference signal [C]//Proceedings of IEEE 86th Vehicular Technology Conference (VTC-Fall). IEEE, 2017: 1–5. DOI: 10.1109/VTCFall.2017.8288402
- [14] POLYANSKIY Y. A perspective on massive random-access [C]//Proceedings of IEEE International Symposium on Information Theory (ISIT). IEEE, 2017: 2523–2527. DOI: 10.1109/ISIT.2017.8006984
- [15] YUAN Y F, WANG S, WU Y P, et al. NOMA for next-generation massive IoT: performance potential and technology directions [J]. IEEE communications magazine, 2021, 59(7): 115–121. DOI: 10.1109/MCOM.001.2000997
- [16] DONOHO D L. Compressed sensing [J]. IEEE transactions on information theory, 2006, 52(4): 1289–1306. DOI: 10.1109/TIT.2006.871582
- [17] CANDES E, TAO T. Near-optimal signal recovery from random projections: universal encoding strategies [EB/OL]. (2004–10–25) [2024–01–10]. <https://arxiv.org/abs/math/0410542>
- [18] DONOHO D L, MALEKI A, MONTANARI A. Message passing algorithms for compressed sensing: II. analysis and validation [C]//Proceedings of IEEE Information Theory Workshop on Information Theory (ITW 2010, Cairo). IEEE, 2010: 1–5. DOI: 10.1109/ITWKSPS.2010.5503228
- [19] MA J J, YUAN X J, PING L. Turbo compressed sensing with partial DFT sensing matrix [J]. IEEE signal processing letters, 2015, 22(2): 158–161. DOI: 10.1109/LSP.2014.2351822
- [20] LIU L, LIANG C L, MA J J, et al. Capacity optimality of AMP in coded systems [J]. IEEE transactions on information theory, 2021, 67(7): 4429–4445. DOI: 10.1109/TIT.2021.3083748
- [21] ORDENTLICH O, POLYANSKIY Y. Low complexity schemes for the random access Gaussian channel [C]//Proceedings of IEEE International Symposium on Information Theory (ISIT). IEEE, 2017: 2528–2532. DOI: 10.1109/ISIT.2017.8006985
- [22] PRADHAN A K, AMALLADINNE V K, VEM A, et al. Sparse IDMA: a joint graph-based coding scheme for unsourced random access [J]. IEEE transactions on communications, 2022, 70(11): 7124–7133. DOI: 10.1109/TCOMM.2022.3183590
- [23] HOWARD S D, CALDERBANK A R, SEARLE S J. A fast reconstruction algorithm for deterministic compressive sensing using second order reed-muller codes [C]//Proceedings of 42nd Annual Conference on Information Sciences and Systems. IEEE, 2008: 11–15. DOI: 10.1109/CISS.2008.4558486
- [24] CALDERBANK R, THOMPSON A. CHIRRAP: a practical algorithm for unsourced multiple access [J]. Information and inference: a journal of the IMA, 2020, 9(4): 875–897. DOI: 10.1093/imaia/iaz029

作者简介



严春林，中关村泛联移动通信技术创新应用研究院资深研究员；长期从事 Beyond 3G、4G、5G、6G 研究工作，主要研究方向为 6G 新型多址、6G 新波形、非正交多址、压缩感知、MIMO 检测、非线性预编码；2014 年独立提出基于 Gray 映射的多用户下行传输方案，该方案是 3GPP MUST 项目的关键技术，2016 年起参与 3GPP 5G NOMA 项目，提出的传输方案达到业界先进水平；发表论文 30 余篇，申请发明专利 20 余项。



袁弋非，中国移动研究院首席专家、IEEE Fellow、海外高层次人才、工信部无线与卫星通信委员会委员、《IEEE Communications Letters》《China Communications》编委、《汉斯-无线通信》期刊主编、3GPP 5G NOMA（非正交多址）Study Item 报告人、IMT-2030（6G）新型多址接入任务组组长；长期从事 3G、4G、5G、6G 的空口关键技术研究 and 标准化工作；获 IEEE 通信协会亚太杰出论文奖，发表论文 60 余篇，出版专著 12 部，获美国授权专利 70 余项。



王森，中国移动研究院无线领域主任研究员、6G 新型无线传输团队负责人，IMT-2030 推进组语义通信任务组副组长；主要研究方向为无线系统优化理论、5G/6G 空口关键技术及标准；发表论文 30 余篇，获发明专利 100 余项。



吕思颖，中国移动研究院研究员；研究方向为 6G 超大规模 MIMO、6G 新型多址技术等；申请专利 10 余项。

基于无线光通信的非正交多址接入技术研究



Non-Orthogonal Multiple Access Technology Based on Optical Wireless Communication

李亮/LI Liang¹, 董宇涵/DONG Yuhuan^{1,2}, 关迅/GUAN Xun^{1,2}, 宋健/SONG Jian^{1,2,3}, 张晓平/ZHANG Xiaoping¹

(1. 清华大学深圳国际研究生院, 中国 深圳 518055;

2. 鹏城实验室, 中国 深圳 518000;

3. 清华大学, 中国 北京 100084)

(1. Tsinghua Shenzhen International Graduate School, Shenzhen 518055, China;

2. Pengcheng Laboratory, Shenzhen 518000, China;

3. Tsinghua University, Beijing 100084, China)

DOI: 10.12142/ZTETJ.202401004

网络出版地址: <http://kns.cnki.net/kcms/detail/34.1228.tn.20240226.1616.013.html>

网络出版日期: 2024-02-27

收稿日期: 2023-12-20

摘要: 将非正交多址接入 (NOMA) 技术应用于无线光通信系统, 可以提高系统容量和频谱效率。首先阐述了基于无线光通信 (OWC) 的非正交多址技术的基本原理和系统结构, 并综述了功率分配、用户分组、波束控制等优化技术的研究进展。之后, 展示了该项技术与无人机 (UAV)、智能反射面 (IRS)、混合光/射频 (RF) 等新兴技术的融合, 这些融合技术可以使能更多的通信场景和需求。最后, 通过使用强化学习来优化一种新型无人机辅助无线光通信非正交多址方案, 并展示了该方案在新型系统中综合性能提升的效果。

关键词: OWC; NOMA; 功率域NOMA; 深度强化学习

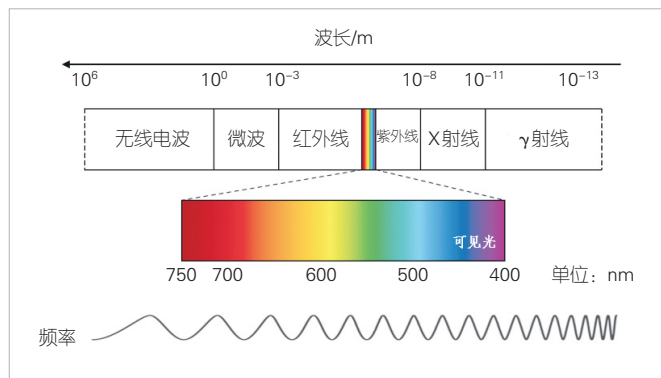
Abstract: Applying non-orthogonal multiple access (NOMA) technology to optical wireless communication (OWC) systems can improve system capacity and spectral efficiency. The basic principle and system structure of OWC-NOMA are first presented, and the related research progress including power allocation, user grouping, beam control, and other optimization techniques is summarized. Then, the integration of OWC-NOMA with emerging technologies such as unmanned aerial vehicles (UAV), intelligent reflecting surface (IRS), and hybrid optical/radio frequency (RF) is introduced, which is capable of facilitating more communication scenarios and requirements. Finally, the overall performance improvement of the system by applying reinforcement learning on a novel UAV-assisted OWC-NOMA scheme is demonstrated.

Keywords: OWC; NOMA; power domain NOMA; deep reinforcement learning

引用格式: 李亮, 董宇涵, 关迅, 等. 基于无线光通信的非正交多址接入技术研究 [J]. 中兴通讯技术, 2024, 30(1): 12-18. DOI: 10.12142/ZTETJ.202401004

Citation: LI L, DONG Y H, GUAN X, et al. Non-orthogonal multiple access technology based on optical wireless communication [J]. ZTE technology journal, 2024, 30(1): 12-18. DOI: 10.12142/ZTETJ.202401004

近年来, 无线光通信 (OWC) 技术受到了学术界和产业界的关注。当前, 射频 (RF) 通信存在频谱资源不足、设备间干扰严重, 以及电磁辐射等问题。OWC具有工作频带宽、免授权频谱、不受电磁干扰、时延低、成本低、部署方便等优点, 将作为6G时代射频通信技术的重要补充^[1]。如图1所示, OWC包括红外、可见光、紫外等频段。目前, OWC技术主要包括可见光通信 (VLC)、自由空间光



▲图1 电磁波谱图

基金项目: 广东省自然科学基金 (2022A1515010209)

通信 (FSO)、紫外通信 (UVC) 等。其中, VLC 可以用于室内、水下、车载等场景, 具有巨大的应用前景; FSO 则凭借高速率远距离的优势, 用于深空通信、数据中心等场景; UVC 基于紫外光子散射, 具有非视距、抗干扰、低窃听等优势, 适用于短距离军用和民用场景。OWC 技术的研究领域涵盖器件、物理层技术、异构组网和高速光互联, 涉及多种技术的创新。可以预见, OWC 技术将在 6G 时代大放异彩, 成为射频通信技术的重要补充, 并与之有机融合以提供更好的服务。

传统的 OWC 系统通常采用正交多址接入 (OMA) 技术, 如频分多址接入 (FDMA)、时分多址接入 (TDMA)、码分多址接入 (CDMA)、正交频分复用多址接入 (OFDMA) 等, 来实现多用户接入。这些技术的基本原理是将信道的频率、时间或码域划分为正交的子信道, 然后分配给不同的用户。然而, 这些技术所能服务的最大用户数量受限于可用正交资源的数量, 并存在信道利用率低、信道分配复杂、同步或滤波要求高等问题。这些问题导致 OMA 技术无法提供足够的资源服务大量用户, 也不能提供大规模的覆盖和连接, 而这却是 6G 网络的重要要求之一^[2]。

为了克服这些缺点, 一种新型的多址技术——非正交多址接入 (NOMA) 被提出。NOMA 技术的基本思想是允许多个用户在同一个时频资源块上同时传输信息, 从而提高信道的利用率和系统的容量。如图 2 所示, 在 1G—5G 时代中, 研究者分别应用 FDMA、TDMA、CDMA、OFDMA 以及大规

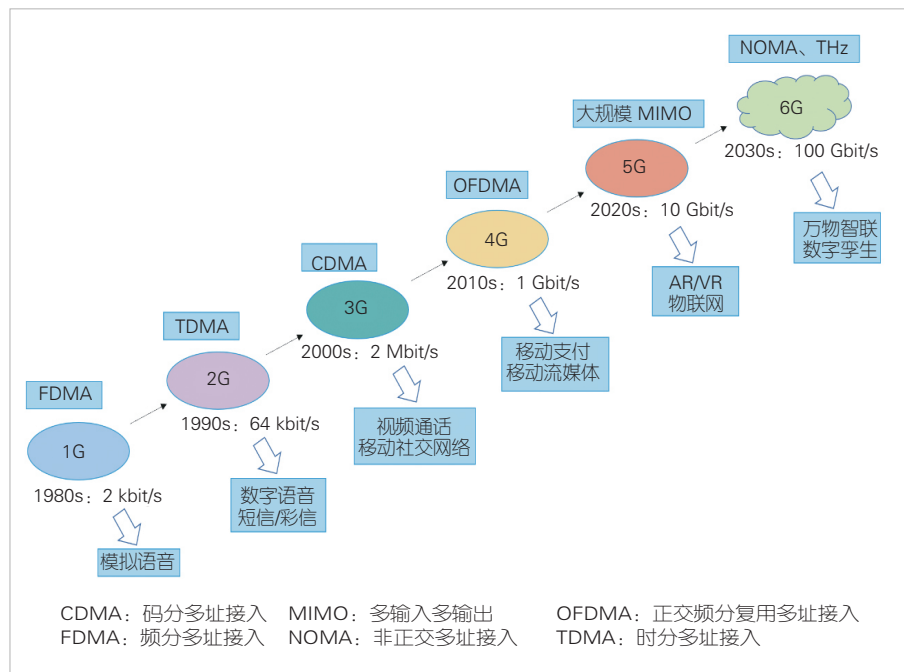
模多输入多输出 (MIMO) 技术来提升通信速率、增加用户接入数量。相对于 4G, 尽管 5G 的性能有了显著改进, 但仍无法满足超低时延、超密集连接、超低功耗的需求。而 NOMA、太赫兹 (THz) 等技术将作为 6G 时代的重要技术, 有效满足复杂场景下低成本、低时延、高吞吐、超密集连接的需求^[1]。NOMA 技术包括功率域非正交多址 (PD)-NOMA 和编码域非正交多址 (CD)-NOMA 两种类型。PD-NOMA 是通过在发射端对不同用户的信号进行不同的功率分配, 然后在接收端利用连续干扰消除 (SIC) 等技术来实现多用户接入的。CD-NOMA 是通过在发射端对不同用户的信号进行不同的编码, 然后在接收端利用联合检测技术来实现多用户接入的^[3]。

基于 OWC 的 NOMA 技术是将 NOMA 技术应用于 OWC 系统的一种通信技术, 可有效地提高 OWC 的频谱效率和系统容量, 同时也可以降低系统的复杂度和同步要求^[2]。PD-NOMA 在 OWC 系统中具有良好的适用性, 主要有以下几个原因: 首先, NOMA 在执行 SIC 解码过程时, 需要准确的信道状态信息 (CSI) 来辅助实现用户复用、解码顺序、功率分配等一系列的功能。在射频通信系统中, 接收端及时获取准确的 CSI 是不容易的, 这是在射频通信中应用 NOMA 的一个主要的限制。但是在 OWC 系统中, 信道状态大多数时间里都是恒定的或者准恒定的, 只会随着用户位置的变化而变化。其次, NOMA 在高信噪比 (SNR) 场景下可以获得更优的性能。在 OWC 系统中, 光源与探测器之间距离较近且多为视距 (LOS) 链路, 因此 SNR 很高。

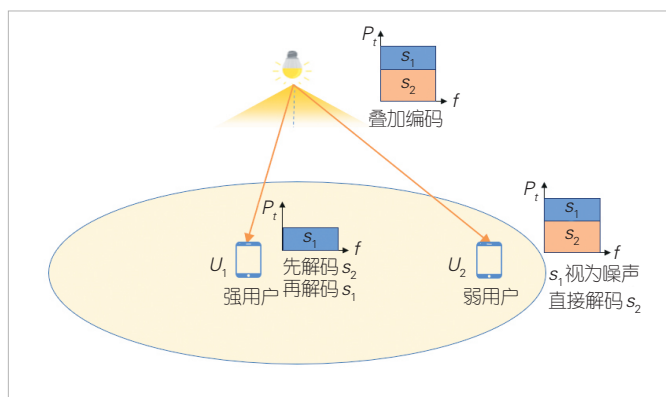
最后, 在 OWC 系统中, 可以动态调整光源和探测器的发射角度和视场角, 增强用户之间信道增益的差异, 从而可以更好地提升 NOMA 的性能。

1 系统结构

PD-NOMA 的主要目标是为不同用户设置不同的功率分配水平。下面我们以一个最简单的情形为例来介绍 PD-NOMA 的系统结构。如图 3 所示, 下行 OWC 基站服务两个用户, 距离较近的用户拥有更好的信道条件, 称为强用户, 而较远的称为弱用户。此时, 基站作为发射端, 为弱用户分配更高功率, 而为强用户分配低功率, 采用叠加编码 (SC) 的方式将这两个用户的期待信号合成, 并将叠加后的信号同



▲图2 移动通信技术的演进:从 1G 到 6G



▲图3 无线光通信-非正交多址接入(OWC-NOMA)系统结构

时发送给这两个用户。而接收时，弱用户最先解码，并将强用户的信号视为噪声。强用户则需要经历SIC的过程，即首先解码弱用户信号并去除，然后解码自己的信号。拓展到多用户情形时，同样是从弱用户到强用户依次解码，每个用户都必须先解码出比自己信道条件弱的用户信号才能获得自己的信号。显然，随着用户数量的增加，解码过程也将变得更加复杂。并且信道估计的误差以及SIC过程中产生的误差，也会累计得更多。因此，我们有必要将NOMA与OMA的技术相结合，并将所有用户分割为多个小组，小组之间采用OMA的技术，小组内则采用NOMA技术。这样的技术方案可以更好地确保小组间不会互相干扰，从而进一步地提升频谱效率。

2 研究进展

NOMA在OWC系统中的应用目前已有大量的研究。早期的研究工作主要集中在通过理论推导和仿真实验来论证在VLC系统中NOMA相较于OMA的优势和性能提升。例如，H. MARSHOUD等^[4]首次将NOMA技术引入VLC系统中，并提出了增益比功率分配（GRPA）的策略，建立了一个完整的NOMA-VLC网络框架。R. C. KIZILIRMAK等^[5]在双用户场景的下行VLC链路中使用NOMA，并与基于OFDMA技术的VLC网络进行比较，仿真结果表明VLC中使用NOMA技术优于OFDMA技术。L. YIN等^[6]针对下行VLC链路推导了系统覆盖概率的解析表达式，并给出了在高SNR条件下，NOMA相较于OMA的总速率增益的上限。作者通过理论和仿真结果证明了NOMA相较于OMA的性能增益，并且可以通过对LED的选择进一步扩大增益，从而实现更好的效果。

随着VLC系统中NOMA技术的应用，研究者将NOMA技术逐步拓展到了FSO以及UVC的非视距链路（NLoS）。例如，L. ZHANG等^[7]采用了垂直腔面发射激光器（VCSEL）实现了基于NOMA的QAM-OFDM的OWC系统，通过调节功率

分配策略，实现两种不同数据流的同时传输，获得1.6 Gbit/s的传输速率。G. C. WANG等^[8]考虑了基于离散时间泊松信道的非视距紫外CDMA和NOMA系统，并推导了可达速率、最佳功率分配的表达式。

近年来，研究者主要研究了功率分配、用户分组、波束控制等技术，进一步优化OWC-NOMA系统的性能。在NOMA系统中，影响系统性能最关键的因素就是对每个用户的功率分配。合理的功率分配是确保SIC解码过程的前提，也会影响系统的吞吐量、用户公平性等。因此，很多研究者提出了不同的功率分配策略，以最大化系统的总速率，最小化误码率，提高用户公平性等。例如，H. MARSHOUD等^[4]提出了GRPA的算法，来确定用户的功率分配系数，从而最大化用户的总速率。

用户分组技术是指在OWC-NOMA系统中，将用户分为几个互不干扰的小组，每个小组之间对应OMA正交的子载波，而在小组内采用NOMA技术。用户分组技术在NOMA中同样至关重要。首先，NOMA在同一资源块中复用的用户数量是有限的，复用过多的用户会导致SIC过程无法执行。其次，有必要在动态的场景中，动态调整用户的分组，简单地将用户分组固定，这样会导致系统性能的下降。因此，很多研究者提出了不同的用户分组策略，以最大化系统的总速率，最小化用户间的干扰，提高用户公平性等为目标。例如，Q. LI等^[9]提出了VLC-NOMA系统中自适应用户关联策略，并比较了无分组和固定分组的性能，通过仿真实验验证了用户分组对性能的影响。

最后，波束控制是OWC-NOMA系统中充分考虑了光源的指向性所产生的技术。随着近年来微机电系统（MEMS）的发展，光源和探测器都可以根据环境动态调整位置和方向，从而在OWC-NOMA系统中跟踪用户，提高信号强度，并在不同VLC接入点之间提供更平滑的切换。基于此，很多研究者提出了不同的波束控制策略，来优化前述的系统总速率、用户间干扰、用户公平性等指标。例如，Y. S. EROGLU等^[10]研究了比例公平速率分配中的最优波束控制参数，特别是用户数量大于波束控制数量的情况。在这种情况下，通过分簇和功率优化，获得的总速率可以达到无转向方案和速率的13倍。

3 新兴技术的融合

3.1 无人机辅助通信

近年来，随着无人机技术的迅猛崛起，通信领域迎来了一场革命性的变革。无人机不再仅是高空俯瞰地表的遥感工

具，它正逐渐演变为多个领域的全方位服务平台。当前基于射频的无人机辅助通信技术存在频率不足、电磁干扰、能耗过高的问题。而将 VLC-NOMA 技术引入无人机辅助通信中，可以在通信的同时提供照明，更适用于救灾的场景，并且具有免许可频谱资源、高速率、低功耗、高安全性、强抗干扰能力的优势。因此，很多学者研究了无人机辅助 VLC-NOMA 系统的性能和优化问题。例如，Q. V. PHAM 等^[11]研究了无人机功率分配和放置的联合问题，通过共同考虑最大功耗约束来最大化总数据速率、用户的服务质量和无人机的位置。他们使用 Harris 鹰优化算法和前馈神经网络来输出无人机的位置和功率分配。A. SULEIMEN 等^[12]采用了一种实验方法来验证无人机辅助 VLC-NOMA 系统的可行性，并通过优化无人机位置来最大化下行链路系统的总容量，表现出良好的误码率和吞吐量性能。

3.2 智能反射面

智能反射面 (IRS) 是一种利用超材料和无源器件实现对电磁波反射的技术，它可以根据不同的通信场景和需求，动态调整反射单元的参数，从而改变信号的幅度、相位等特性，实现信号波束成形的控制。IRS 具有低成本、低复杂度、高灵活性、高性能等优势，可以用于解决信号盲区，提升通信安全，提高通信质量。

传统的 OWC 系统在面临多用户接入的场景时，往往存在个别用户由于障碍物遮挡处在光信号的盲区，从而无法建立直射链路的问题。将 IRS 应用于 OWC-NOMA 系统中，可以制造出反射路径，服务光信号盲区的用户，完成多用户接入的需求。例如，Y. H. JIAN^[13]等提出将空间光调制器 (SLM) 作为 IRS 进行光束控制，通过理论和实验分析了 SLM 在担任 IRS 时的一些特点。基于 SLM 的 IRS 可以独立同时控制多个波束，从而实现多用户的接入需求。P. SINGH^[14]等通过在无人机上搭载 IRS 解决基站与移动用户之间的跳区问题，并设计了基于 SNR 的镜像分配策略，将 IRS 的多个镜像单元分组并分配给用户，从而实现了每个用户的最大吞吐量和比例公平。

3.3 混合光/射频技术

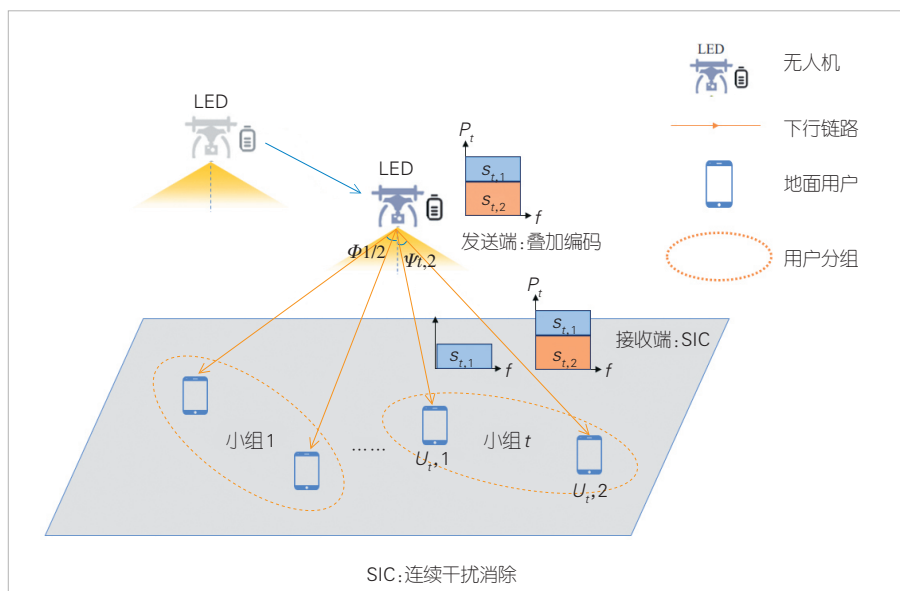
在 6G 的愿景中，OWC 并非要取代射频通信，而是作为射频通信的重要补

充。OWC 在下行直射链路中展示了其高速率、低延迟、低能耗的优势；但是当面临非直射链路以及上行链路时，与射频通信系统互补是一个很好的选择。因此，开发出混合 OWC/RF 技术，从而融合 OWC 和 RF 各自的优势很有必要。M. B. RAHAIM^[15]等已经证明了混合 RF/VLC 网络可以显著提升能量效率，扩大服务范围，提高总系统容量，可以作为室内通信网络的优先选择。将 NOMA 技术引入 RF/OWC 系统同样是有利的。D. Y. SHI^[16]等提出了一种与 NOMA 相结合的混合 RF-VLC 链路，支持多用户同时从服务器获取数据，以及从本地物联网设备中获取传感器数据。由于 RF 与 VLC 信道之间不存在干扰，以及 NOMA 可以分配给用户更多的时频资源，这些因素有效提高了通信性能并降低了误码率。

4 基于强化学习的无人机辅助 VLC-NOMA

如图4所示，考虑一个无人机辅助 VLC 系统。无人机通过搭载 LED 可以同时实现通信和照明的服务，以应对灾后救援、流量卸载等场景。为应对当前商用 LED 中调制带宽不足的问题，考虑引入 NOMA 技术来提高系统的频谱效率和用户接入的数量。

在无人机辅助 VLC-NOMA 系统中，需要联合优化无人机的轨迹、用户分组和功率分配从而提升系统性能。这是一个非凸且 NP-hard 的问题，传统的优化算法难以求解。强化学习是机器学习的一个重要分支，可通过与环境交互获得奖励，进而指导智能体学习最优决策，达到累计奖励最大化的目标，非常适用于应对无人机辅助通信中轨迹规划的序列决



▲图4 无人机辅助可见光通信-非正交多址(VLC-NOMA)系统

策问题。

4.1 系统模型和问题表述

对于通信信道，考虑无人机和地面用户之间的 LoS 链路，无人机和用户 $U_{t,k}$ 之间的信道增益可以表示为^[11]：

$$h_{t,k} = \begin{cases} \frac{A(m+1)}{2\pi d_{t,k}^2} g(\psi_{t,k}) T_s(\phi_{t,k}) \cos^m(\phi_{t,k}) \cos(\psi_{t,k}) \\ 0 & 0 < \psi_{t,k} < \Psi_c \\ 0, & \psi_{t,k} > \Psi_c \end{cases}, \quad (1)$$

其中， $\psi_{t,k}$ 、 $\phi_{t,k}$ 和 Ψ_c 分别是入射角、辐照度角、接收器视场的半角， A 是探测器面积， $d_{t,k}$ 是无人机到用户 $U_{t,k}$ 的距离， $T_s(\phi_{t,k})$ 为滤光片增益。 $g(\psi_{t,k})$ 表示光学集中器的增益，当 $0 \leq \psi_{t,k} \leq \Psi_c$ 时， $g(\psi_{t,k}) = \frac{n_r^2}{\sin^2 \Psi_c}$ ，否则为 0。 m 与发射器半发散角 $\Phi_{1/2}$ 有关，并且 $m = -\ln 2 / \ln(\cos \Phi_{1/2})$ 。

我们将总带宽 B 平均划分为 T 个子信道，对应不同无干扰小组。小组内采用 NOMA 方式实现时频资源共享。考虑 NOMA 中最简单的情况，即每小组 2 个用户，并假设 $h_{t,1} \geq h_{t,2}$ 。根据 NOMA 的规则，功率分配系数 $a_{t,1} \leq a_{t,2}$ ，且 $a_{t,1} + a_{t,2} = 1$ ，可以得出 $a_{t,1} \leq 0.5$ 。则小组 t 中所有用户的下行可达速率可以表示为：

$$R_t = \underbrace{\log_2 \left(1 + \frac{h_{t,1} a_{t,1} P_t}{n_0} \right)}_{R_{t,1}} + \underbrace{\log_2 \left(1 + \frac{h_{t,2} a_{t,2} P_t}{n_0 + h_{t,2} a_{t,1} P_t} \right)}_{R_{t,2}}. \quad (2)$$

在给定功率 $P_{\max}$ 时，以最大化所有用户的总速率为目标，优化问题可以数学表示为：

$$\begin{aligned} & \max_{w, T, \alpha} \sum_t R_t \\ & \text{s.t. } h_{t,1} \geq h_{t,2} \\ & \quad a_{t,1} \leq 0.5 \\ & \quad R_{t,k} \geq R_{th} \\ & \quad w \in \mathcal{D} \\ & \quad P_t \geq 0 \\ & \sum_{t \in T} P_t \leq P_{\max}, \end{aligned} \quad (3)$$

其中， $w = (x_u, y_u, z_u)$ 是无人机位置， T 为用户分组策略， $\alpha = \{a_{1,1}, a_{1,2}, \dots, a_{T,1}, a_{T,2}\}$ 为功率分配系数。优化 $\{w, T, \alpha\}$ 受到如下多个约束：NOMA 功率分配规则； R_{th} 是所有用户满足服务质量 (QoS) 条件的最小速率要求；无人机限制在 $\mathcal{D}$ 区域内飞行；每小组 t 的功率非负且总和不超过总功率 $P_{\max}$ 。该问

题非凸且 NP-hard，传统优化算法很难求解。此外，VLC 信道模型形式也导致传统凸近似算法不适合求解。因此，我们采用深度强化学习来解决该问题。

4.2 深度强化学习算法

在强化学习中，环境通常被建模为马尔科夫决策过程，包括状态空间、动作空间、奖励、转移概率分布和折扣因子。公式 (3) 中的优化问题可转化为强化学习环境。

状态空间包括无人机位置、用户的信道增益、用户通信速率、总通信速率，可表示为： $s = [w, h_{1,1}, h_{1,2}, \dots, h_{T,1}, h_{T,2}, R_{1,1}, R_{1,2}, \dots, R_{T,1}, R_{T,2}, \sum_{t \in T} R_t]$ 。

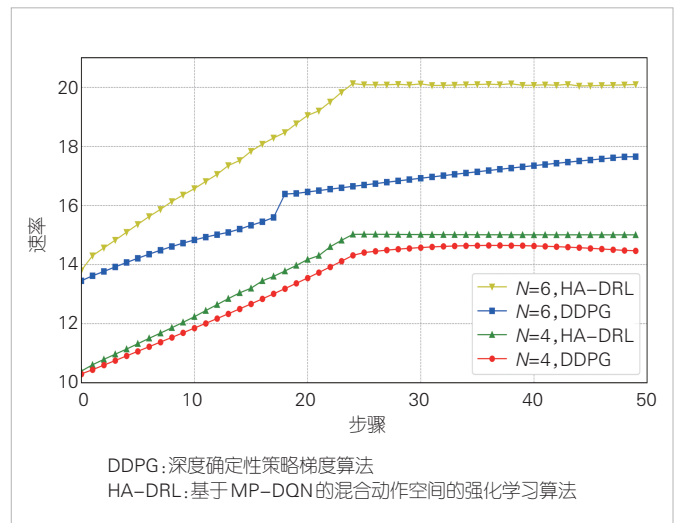
动作空间包括离散和连续两种类型，离散动作包括用户分组，所有分组的可能性为 K ；连续动作包括无人机位移和功率分配系数，可表示为 $x_k = (v_x, v_y, v_z, a_1, a_2, \dots, a_T)$ 。

奖励函数与总通信速率正相关，可表示为 $\mathcal{R} = \sum_{t \in T} R_t - \mathcal{C}$ ，其中， $\mathcal{C}$ 为一个常数，用来控制奖励值在一个合理的范围。

在算法层面，我们提出了基于多通道深度 Q 网络 (MP-DQN)^[17] 的混合动作空间的强化学习 (HA-DRL) 算法来分别处理离散和连续的动作。为了体现混合动作空间的优势，我们还与单一连续动作空间的深度确定性策略梯度算法 (DDPG)^[18] 进行比较。

4.3 结果与分析

经过 2 500 个回合的训练后，我们将保存 HA-DRL 和 DDPG 的模型参数并测试。测试结果如图 5 所示，可以看出，当用户数量由 4 增加到 6 后，HA-DRL 相较于 DDPG 算法的



▲图5 不同强化学习算法的测试曲线比较

优势进一步增加。这是因为随着用户数量增加,映射到离散动作所需要的连续动作也更多,动作空间增加的也更多。此时传统DDPG算法更难探索到更好的策略,训练也更难收敛,相应的测试结果也会更差。

5 结束语

本文中,我们介绍了OWC-NOMA技术的基本原理、系统结构、研究进展和新兴技术的融合,展示了OWC-NOMA技术在6G时代的巨大应用潜力和应用前景。OWC-NOMA技术可以有效地提高OWC的频谱效率和系统容量,同时也可以降低系统的复杂度和同步/滤波要求。OWC-NOMA技术还可以与无人机、智能反射面、混合光/射频等一系列新兴的技术相结合,服务更多的通信场景和需求。最后,介绍了一项无人机辅助VLC-NOMA的研究进展,并通过深度强化学习方法优化了系统的性能。OWC-NOMA技术是一种具有创新性和前瞻性的通信技术,值得我们进行进一步的研究和探索。

参考文献

- [1] WANG C X, YOU X H, GAO X Q, et al. On the road to 6G: visions, requirements, key technologies, and testbeds [J]. IEEE communications surveys & tutorials, 2023, 25(2): 905–974. DOI: 10.1109/COMST.2023.3249835
- [2] MOHSAN S A H, SADIQ M, LI Y, et al. NOMA-based VLC systems: a comprehensive review [J]. Sensors (basel), 2023, 23(6): 2960. DOI: 10.3390/s23062960
- [3] DAI L L, WANG B C, DING Z G, et al. A survey of non-orthogonal multiple access for 5G [J]. IEEE communications surveys & tutorials, 2018, 20(3): 2294–2323. DOI: 10.1109/comst.2018.2835558
- [4] MARSHOUD H, KAPINAS V M, KARAGIANNIDIS G K, et al. Non-orthogonal multiple access for visible light communications [J]. IEEE photonics technology letters, 2016, 28(1): 51–54. DOI: 10.1109/pt.2015.2479600
- [5] KIZILIRMAK R C, ROWELL C R, UYSAL M. Non-orthogonal multiple access (NOMA) for indoor visible light communications [C]//Proceedings of 4th International Workshop on Optical Wireless Communications (IWOW). IEEE, 2015: 98–101. DOI: 10.1109/iwow.2015.7342274
- [6] YIN L, POPOOLA W O, WU X P, et al. Performance evaluation of non-orthogonal multiple access in visible light communication [J]. IEEE transactions on communications, 2016, 64(12): 5162–5175. DOI: 10.1109/tcomm.2016.2612195
- [7] ZHANG L, WEI Z X, CHEN Y J, et al. VCSEL-based multi-user optical wireless communication system using non-orthogonal multiple access [C]//Proceedings of 14th Pacific Rim Conference on Lasers and Electro-Optics (CLEO PR 2020). IEEE, 2020. DOI: 10.1364/cleopr.2020.c3f_5
- [8] WANG G C, GONG C, XU Z Y. Signal characterization for multiple access non-line of sight scattering communication [J]. IEEE transactions on communications, 2018, 66(9): 4138–4154. DOI: 10.1109/tcomm.2018.2830766
- [9] LI Q, SHANG T, TANG T, et al. Adaptive user association scheme for indoor multi-user NOMA-VLC systems [J]. IEEE wireless communications letters, 2023, 12(5): 873–877. DOI: 10.1109/lwc.2023.3247420
- [10] EROGLU Y S, ANJINAPPA C K, GUVENC I, et al. Slow beam steering and NOMA for indoor multi-user visible light communications [J]. IEEE transactions on mobile computing, 2021, 20(4): 1627–1641. DOI: 10.1109/tmc.2019.2960495
- [11] PHAM Q V, HUYNH-THE T, ALAZAB M, et al. Sum-rate maximization for UAV-assisted visible light communications using NOMA: swarm intelligence meets machine learning [J]. IEEE Internet of Things journal, 2020, 7(10): 10375–10387. DOI: 10.1109/jiot.2020.2988930
- [12] SULEIMEN A, KIZILIRMAK R C. UAV assisted vehicular communication with VLC using NOMA [C]//Proceedings of 12th International Symposium on Communication Systems, Networks and Digital Signal Processing (CSNDSP). IEEE, 2020: 1–5. DOI: 10.1109/csndsp49049.2020.9249511
- [13] JIAN Y H, WANG C C, CHOW C W, et al. Optical beam steerable orthogonal frequency division multiplexing (OFDM) non-orthogonal multiple access (NOMA) visible light communication using spatial-light modulator based reconfigurable intelligent surface [J]. IEEE photonics journal, 2023, 15(4): 1–8. DOI: 10.1109/jphot.2023.3294834
- [14] SINGH P, BOHARA V A, SRIVASTAVA A. On the optimal assignment of mirror element in UAV and OIRS-assisted OWC based architecture [C]//Proceedings of IEEE 97th Vehicular Technology Conference (VTC2023-Spring). IEEE, 2023: 1–7. DOI: 10.1109/vtc2023-spring57618.2023.10199990
- [15] RAHAIM M B, VEGNI A M, LITTLE T D C. A hybrid Radio Frequency and broadcast Visible Light Communication system [C]//Proceedings of IEEE GLOBECOM Workshops (GC Wkshps). IEEE, 2011: 792–796. DOI: 10.1109/glocomw.2011.6162563
- [16] SHI D Y, ZHANG X, SHI L N. A joint backscatter and VLC-NOMA communication scheme for B5G/6G umMTC system [C]//Proceedings of IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB). IEEE, 2021: 1–4. DOI: 10.1109/bmsb53066.2021.9547027
- [17] BESTER C J, JAMES S D, KONIDARIS G. Multi-pass Q-networks for deep reinforcement learning with parameterised action spaces [EB/OL]. (2019-05-10)[2024-01-04]. <https://arxiv.org/abs/1905.04388>
- [18] LILLICRAP T, HUNT J J, PRITZEL A, et al. Continuous control with deep reinforcement learning [EB/OL]. (2019-07-05)[2024-01-04]. <https://arxiv.org/abs/1509.02971>

作者简介



李亮, 清华大学深圳国际研究生院在读硕士生; 主要研究方向为无人机辅助无线光通信, 具体包括无人机路径规划、深度强化学习等。



董宇涵，清华大学深圳国际研究生院副教授、数据与信息研究院院长助理，IEEE Senior Member、中国光学工程学会高级会员；长期从事无线通信与网络、机器学习与优化、智能传感器网络及应用、人工智能与医疗、智能交通车路协同等方面的研究；作为项目（技术）负责人承担国家自然科学基金、国家重点研发计划子课题、省部级项目、企业委托项目30余项；发表论文200余篇，申请专利60余项，授权专利40余项。



关迅，清华大学深圳国际研究生院助理教授、加拿大拉瓦尔大学客座教授；主要研究方向包括光通信、光电集成、通信系统与信号处理；担任《IEEE JSAC》《Journal of Lightwave Technology》《IEEE Photonics Technology Letters》等国际期刊审稿人和IPC、ONDM等多个国际会议的程序委员会成员；主持和参与多项各级科研项目；发表论文50余篇。



宋健，清华大学电子工程系教授、清华大学深圳国际研究生院数据与信息研究院副院长，IEEE Fellow，IET Fellow，中国电子学会会士，中国通信学会会士，国际电联《智能融合网络》创始主编，《IEEE Transaction on Broadcasting》副主编，《IEEE Access for BTS》主编，现任清华大学数字电视技术研究中心主任，数字电视国家工程实验室（北京）主任；主要研究方向为无线数字广播、网络融合、无线通信、可见光通信与电力线通信等；发表论文300余篇，授权专利80余项，出版中英文专著6部。



张晓平，清华大学鹏瑞教授、清华大学深圳国际研究生院数据与信息研究院院长，加拿大工程院院士，加拿大工程研究院院士，IEEE Fellow，清华-伯克利深圳学院（TBSI）信息和数据科学首席教授，《IEEE 信号处理专刊》主编，IEEE 信号处理学会图像视频及多维信号处理技术委员会主席，IEEE 信号处理学会杰出讲座学者以及IEEE 电路和系统学会杰出讲座学者；研究方向为信号处理和大数据的理论及应用，主要从事统计模型、信号处理、机器学习和人工智能、物联网和电子信息系统、多媒体、生物信息及金融经济模型和大数据等方面的研究。

面向6G移动通信的极化多址接入技术



Polar Coded Multiple Access Technology for 6G Mobile Communication

崔宏基/CUI Hongji, 牛凯/NIU Kai

(北京邮电大学泛网无线通信教育部重点实验室, 中国 北京 100876)
(The Key Laboratory of Universal Wireless Communications, Ministry of Education, Beijing University of Posts and Telecommunications, Beijing 100876, China)

DOI: 10.12142/ZTETJ.202401005

网络出版地址: <http://kns.cnki.net/kcms/detail/34.1228.TN.20240228.0936.002.html>

网络出版日期: 2024-02-27

收稿日期: 2023-12-16

摘要: 极化码是第一种达到信道容量的编码方案, 是5G信道编码方案之一。面对6G更高可靠度、更低时延、更大连接的业务需求, 提出面向6G的极化多址接入技术。极化多址接入技术能够充分利用多用户之间的极化效应, 以实现更高的系统容量和系统接入能力。有源极化编码接入方案能够满足6G大容量高可靠的传输需求, 无源极化编码随机接入方案能够提高系统的接入能力。这些方案表明, 极化多址接入技术能够满足6G移动通信系统的超高可靠、高频谱效率、大容量的传输需求, 是重要的候选技术之一。

关键词: 多址接入; 极化编码; 随机接入; 非正交多址接入

Abstract: Polar codes are the first encoding scheme to achieve channel capacity and are employed in one of the channel coding schemes in 5G. Faced with the demands of 6G for higher reliability, lower latency, and increased connectivity, polar-coded multiple access techniques tailored for 6G are proposed. The polar-coded multiple access technique leverages the polarization effect among multiple users to achieve higher system capacity and access capability. The sourced polar-coded multiple access scheme can meet the high-capacity and high-reliability transmission requirements of 6G, while the unsourced polar-coded random access scheme enhances the system's access capability. These approaches indicate that polar-coded multiple access technologies can address the ultra-high reliability, high spectral efficiency, and large-capacity transmission requirements of 6G mobile communication systems, making them a crucial candidate technology.

Keywords: multiple access; polar code; random access; non-orthogonal multiple access

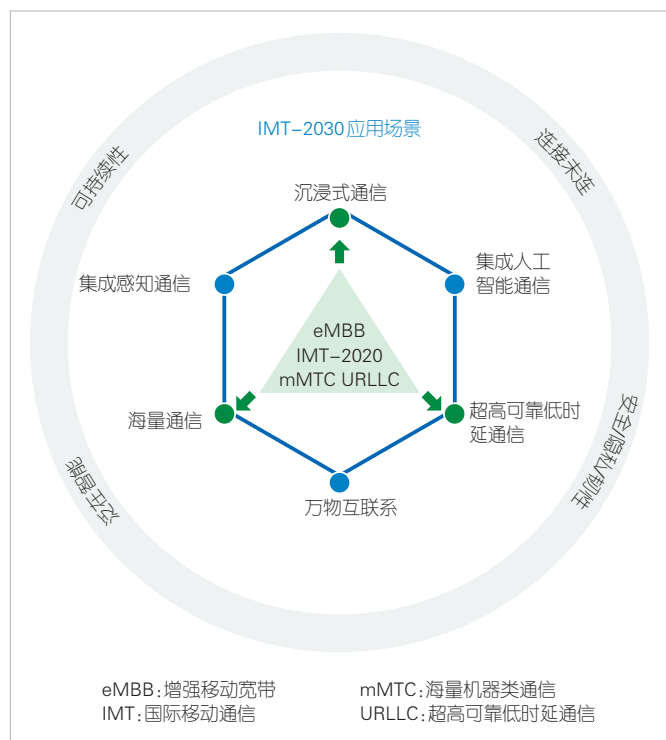
引用格式: 崔宏基, 牛凯. 面向6G移动通信的极化多址接入技术[J]. 中兴通讯技术, 2024, 30(1): 19-25. DOI: 10.12142/ZTETJ.202401005

Citation: CUI H J, NIU K. Polar coded multiple access technology for 6G mobile communication [J]. ZTE technology journal, 2024, 30(1): 19-25. DOI: 10.12142/ZTETJ.202401005

随着6G标准化逐渐提上日程, 全球移动通信技术的争夺焦点正迅速转向6G。2023年6月, 国际电信联盟无线电通信部门 (ITU-R) 的新建议草案^[1]提出了6G通信的六大应用场景: 沉浸式通信、海量通信、超高可靠度低延迟通信、万物互联、集成人工智能通信是和集成感知通信, 具体如图1所示。相比于5G移动通信, 6G对接入数量、接入类型、可靠度和时延都提出更为严格的要求。同时, 6G通信技术以更高级别“智能”促进网络“简约”, 驱动网络动态重构形成“智简无线网络”。本文所提出的极化多址技术是满足大连接高可靠通信的关键技术途径。

多址接入技术旨在满足多个用户的接入需求, 提高系统频谱效率。多址接入技术经历了1G的频分多址接入 (FDMA)、2G的时分多址接入 (TDMA)、3G的码分多址接入 (CDMA)、4G和5G的正交频分多址接入 (OFDMA)。这些多址技术都是正交多址接入 (OMA) 技术, 对多用户信号进行正交划分。虽然这些技术的信号处理复杂度较低, 但并不能达到上下行信道容量域。非正交多址接入 (NOMA) 技术是6G空口技术的关键技术, 其核心思想是突破正交多址接入时域、频域或者码域资源的正交划分, 多个用户信号在相同资源上叠加传输, 从而显著提升频谱效率。2017年, Y. POLYANSKIY提出无源多址接入 (UMA)^[2]的概念。不同于传统OMA技术, UMA移除了网络的协作中心来以降低频繁访问大量短数据包的传输成本和延迟, 是一种无协调免授

基金项目: 国家自然科学基金 (62071058)

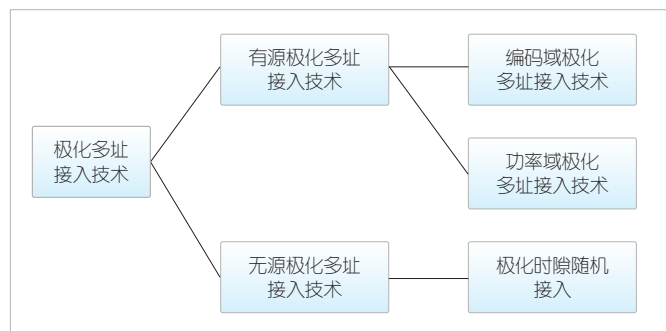


▲图1 IMT-2030 六大使用场景^[1]

权大规模随机接入方案。

2009 年，土耳其学者 E. ARIKAN 在文献[3]中基于信道极化思想，提出了极化码，首次以构造性方法证明信道容量渐近可达。极化码自发明 10 年来，已成为信道编码领域的热门研究方向，其理论基础已经初步建立。2016 年底，极化码入选 5G 移动通信的控制信道编码候选方案，并最终写入 5G 标准。6G 移动通信网络面临繁多、复杂的应用场景，极化编码多址接入技术以“智简”的方式应对多样化场景带来的高可靠、高频谱效率、大连接等技术挑战，将是很有竞争力的一种候选技术。

本文按照发送消息是否含有身份标识，将极化多址接入技术分为两大类（如图 2 所示）：以提升系统容量为目标的有源多址接入和以提高接入能力为目标的无源多址接入，分



▲图2 本文涉及的极化多址技术分类

别对应高可靠、大连接两种典型场景。

在有源多址接入中，协作中心负责管理用户接入的碰撞行为，并且通过插入一定的信息，甄别接入用户的身份，由此在接收端可知可感用户的碰撞模式。这种方式有利于功率域叠加多址接入与编码域多址接入技术的实现。

而无源多址接入技术无须协作中心介入，用户接入会发生随机碰撞，同时不须要传输协作信息，没有协作资源的额外开销，因此接入处理时延更低。无源多址接入无统一接入调度，无导频和预分配资源，用户活跃状态未知。由于无协作中心参与，每个资源上承载用户数量随机，不能事先确定用户的码本向量。所有用户使用完全相同的一套传输协议，随机占用接入资源，系统是完全开环、单向的信号传输。

本文介绍了极化编码的基本原理，分析有源极化多址接入和无源极化多址接入的系统架构，并分析极化编码在有源多址接入下提升频谱效率的基本思想，提出无源极化多址接入大连接的接入方案。

1 极化编码原理

极化编码方案是指利用极化现象来进行的信道编码方案。信道极化由 E. ARIKAN 在文献[3]引入，是指对一组可靠性相同的二进制对称输入离散无记忆信道进行信道合并和信道拆分，变换为一组有相关性的、可靠性各不相同的极化子信道的过程。随着码长的增加，这些子信道可靠性呈现两极分化现象。ARIKAN 证明，当信道数目充分大时，极化子信道的互信息完全两极分化为：无噪的好信道（互信息趋于 1）与完全噪声的差信道（互信息趋于 0），并且好信道占总信道的比例趋于原始二进制对称输入离散无记忆信道的容量 $I(W)$ ，而差信道比例趋于 $1-I(W)$ 。

在极化编码方案中，我们首先需要对信道的可靠性进行评估，选择出当前 N 个子信道中可靠度最高的 K 个子信道，选定为信息位集合 $\mathcal{A}$ ，剩余 $N-K$ 个子信道用作冻结比特集合 $\mathcal{A}^c$ 。冻结比特对编码器和译码器都是已知的，一般可以固定为零。子信道的选取有多种构造方式，可以根据不同的信道、译码算法和构造复杂度进行选取，详细的构造方式可参阅文献[4]。下面我们给出一般性的极化编码方案。

给定 (N, K) 极化码，其中信息位长度为 K ，编码长度为 N 。编码器输入比特序列由信息比特与冻结比特构成，表示为 $u_1^N = (u_1, u_2, \dots, u_N) = (u_{\mathcal{A}}, u_{\mathcal{A}^c})$ 。令 $x_1^N = (x_1, x_2, \dots, x_N)$ 表示编码比特序列，则极化码的编码表示为：

$$x_1^N = u_1^N G_N, \quad (1)$$

其中，编码生成矩阵 $G_N = B_N F_N^{\otimes n}$ ， B_N 是排序矩阵，表示比

特反序操作, $\mathbf{F}^{\otimes n}$ 表示下三角矩阵 $\mathbf{F} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$ 进行 n 次克罗内克积操作得到的 Hadamard 矩阵。由于极化码采用蝶形结构编码, 因此极化码的编码复杂度为 $O(M \log N)$ 。

目前极化码译码有两类算法: 一类是基于串行抵消 (SC) 译码算法, 另一类是基于置信度传播 (BP) 译码算法。循环冗余校验码辅助的串行抵消列表/堆栈译码算法 (CA-SCL/SCS) [5] 是提升有限码长下极化码性能的关键, 它可以使有限码长下极化码性能超越低密度奇偶校验码 (LDPC) / Turbo 码。业界提出的自适应 CA-SCL 算法 [6], 极大增强了译码性能, 并且平均复杂度降低 5~10 倍, 成为极化码应用于 5G 移动通信的关键优势算法。循环冗余校验码辅助的极化码混合球译码算法 [7] 在码长为 128 时仅与有限码长容量限相差 0.025 dB。BP 类译码算法在性能上稍逊, 但具有良好的并行结构, 是高吞吐率译码器设计的关键。

2 有源极化多址接入技术

NOMA 是一类提高多用户通信系统容量的有效方法, 属于有源多址接入技术。其中, 稀疏码多址接入 (SCMA)、图样分割多址接入 (PDMA) 等方案, 能够为每个用户分配不同的码本向量。在占用同等时频资源条件下, 相比于正交多址接入, NOMA 能够增加接入用户数量。由于 NOMA 系统各个用户的检测可靠性存在差异, 因此也可以看作是广义极化。采用极化编码的多址接入方案, 是满足 6G 大容量接入的重要候选技术。

2.1 编码域极化多址接入

文献 [8] 提出了极化编码 NOMA (PC-NOMA) 的基本框架, 如图 3 所示。PC-NOMA 系统针对的是多址接入信道, 会在发送端每个用户的数据分别进行极化编码、交织、星座调制与多用户码本映射, 然后送入信道。在接收端, 接收信号首先在多用户因子图上进行软入软出检测, 产生每一路用户数据的软信息, 然后送入软解调单元, 得到比特似然比信息, 最后送入各个用户的极化码译码器进行纠错。

PC-NOMA 系统的极化变换可以划分为三级极化: 用户极化、调制极化和编码极化。其中, 编码极化利用了编码比特之间的可靠性差异, 调制极化利用了星座比特的可靠性差异, 而用户极化充分利用了用户可靠性之间的差异。我们将 NOMA 的多用户码本映射过程看作是广义的极化变换, 采用极化编码调制进行充分匹配, 从而获得系统的整体优化。

PC-NOMA 系统的检测结构如图 3 所示, 多用户检测算

法按照一定的顺序输出某个用户的软信息, 进行软解调、解交织与极化译码后, 将判决结果反馈到多用户检测单元, 进行干扰抵消后, 再对下一个用户继续进行检测、解调与译码。

对于上行 NOMA 系统, 假设有 J 个用户, 第 v 个用户的码本向量为 $\mathbf{x}_v = (x_{v,1}, x_{v,2}, \dots, x_{v,F})$, 对应的信道衰落向量为 $\mathbf{h}_v = (h_{v,1}, h_{v,2}, \dots, h_{v,F})$ 。则接收信号向量 $\mathbf{y} = (y_1, y_2, \dots, y_F)$ 可以表示为:

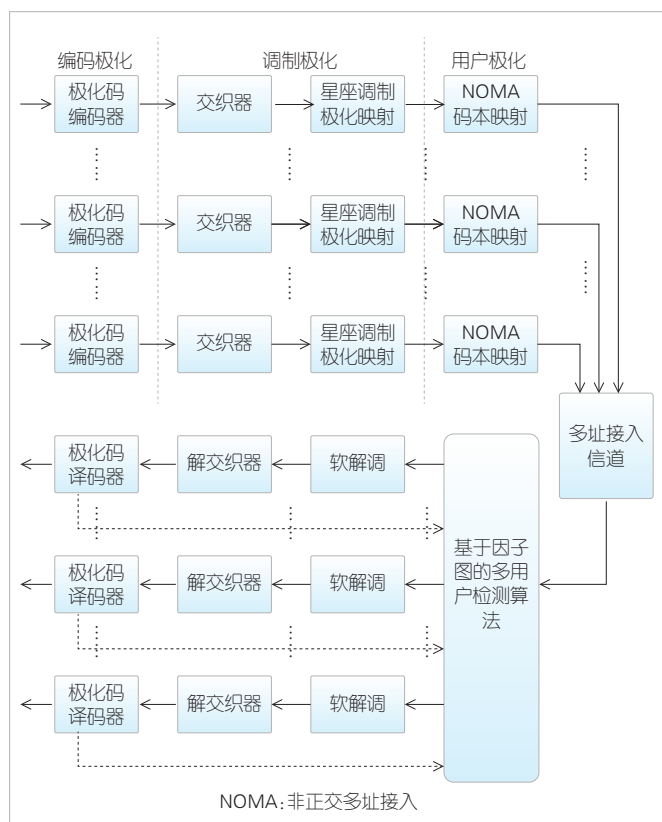
$$\mathbf{y}^T = \sum_{v=1}^J \text{diag}(\mathbf{h}_v) \mathbf{x}_v^T + \mathbf{z}^T, \quad (2)$$

其中, $()^T$ 表示向量转置, $\text{diag}(\mathbf{h}_v)$ 表示以向量 $\mathbf{h}_v$ 的衰落系数为对角线元素的对角矩阵, $\mathbf{z}^T$ 为加性噪声向量。

假设每个用户采用 2^m 进制调制, 发送 N 个符号, 根据文献 [8] 的分析, 利用互信息链式法则, PC-NOMA 信道的序列互信息表示为:

$$I(b_1, b_2, \dots, b_J; \mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_N) = \sum_{v=1}^J \sum_{l=1}^m \sum_{i=1}^N I(w_{v,j,i}), \quad (3)$$

其中, $\mathbf{b}_v$ 表示第 v 个用户发送的信息序列, $\mathbf{y}_l$ 表示第 l 个时刻接收到的信号向量, $w_{v,j,i}$ 表示经过用户/调制/编码三级极化



▲图3 极化编码NOMA系统结构

后,对应的第 v 个用户、第 j 个比特、第 i 个极化子信道,其互信息为 $I(W_{v,j,i})$ 。

为了考察不同编码方式的NOMA系统性能,我们仿真了高斯白噪声(AWGN)信道下,极化码稀疏码多址(PC-SCMA)、极化码图样分割多址接入(PC-PDMA)、Turbo编码稀疏码多址(TC-SCMA)与Turbo编码图样分割多址接入(TC-PDMA)系统的性能。这里,“PC”表示极化编码,“TC”表示Turbo编码。每个用户的极化码或者Turbo码长为 $N=1024$,所有用户的平均码率 $R=1/2$ 。极化码采用CA-SCL译码算法,Turbo码采用对数最大后验估计(Log-MAP)译码算法。

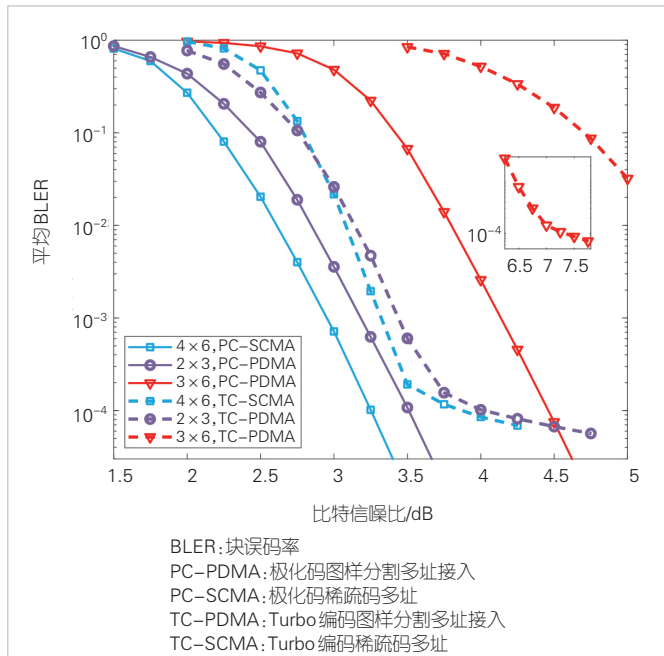
4×6 SCMA 码本矩阵、 2×3 与 3×6 PDMA 码本矩阵分别定义如下:

$$\mathbf{F}_{4 \times 6}^{\text{SCMA}} = \begin{bmatrix} 0 & 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 1 & 0 \end{bmatrix}, \mathbf{F}_{2 \times 3}^{\text{PDMA}} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix},$$

$$\mathbf{F}_{3 \times 6}^{\text{PDMA}} = \begin{bmatrix} 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 \end{bmatrix}, \quad (4)$$

其中,SCMA码本对应的负载为150%,PDMA码本相应的负载分别为150%与200%。

AWGN信道下,4种配置的块误码率(BLER)性能如图4所示。从图中可以看出,无论是采用PDMA码本还是



▲图4 高斯白噪声(AWGN)信道下不同配置的非正交多址接入(NOMA)方案BLER性能比较^[8]

SCMA码本,极化编码系统都比Turbo编码系统有显著的性能增益。例如,采用 3×6 PDMA码本,当 $\text{BLER}=10^{-4}$ 时,PC-PDMA相比于TC-PDMA可以获得3 dB的性能增益。并且,TC-SCMA或TC-PDMA都出现了明显的错误平台现象,而PC-SCMA或PC-PDMA都没有该现象。

通过上述比较,可以看出,极化编码NOMA系统是一种提高6G系统容量的有效手段,是具有竞争力的候选方案。

2.2 功率域极化多址接入

不同于编码域NOMA技术,功率域极化多址技术使用功率来区分用户。在进行功率分配的过程中,两用户的容量不是相同的,将功率极化效应定义为两级极化变换。对于两用户信道 W_1 、 W_2 ,考虑功率分配和极化编码的级联变换。

定义第一级功率变换为:

$$\psi_1: (W_1, W_2) \mapsto (W'_1, W'_2). \quad (5)$$

我们考虑在信道变换中信道互信息的均值和方差变化:

$$M_{\psi_1} = \frac{1}{2} \sum_{k=1}^2 C_k = \frac{1}{2} (C_1 + C_2),$$

$$V_{\psi_1} = \frac{1}{2} \sum_{k=1}^2 C_k^2 - M_{\psi_1}^2 = \left(\frac{1}{2} \cdot |C_1 - C_2| \right)^2, \quad (6)$$

其中, C_1 、 C_2 分别是信道 W_1 、 W_2 的信道容量。很显然, C_1 、 C_2 受到功率分配的影响。

第二级变换时极化编码造成的比特子信道的极化:

$$\psi_2: (W'_1, W'_2) \mapsto (\{W_N^{(1,1)}, \dots, W_N^{(1,N)}\}, \{W_N^{(2,1)}, \dots, W_N^{(2,N)}\}). \quad (7)$$

对于整个级联系统,其极化子信道互信息的均值和方差为:

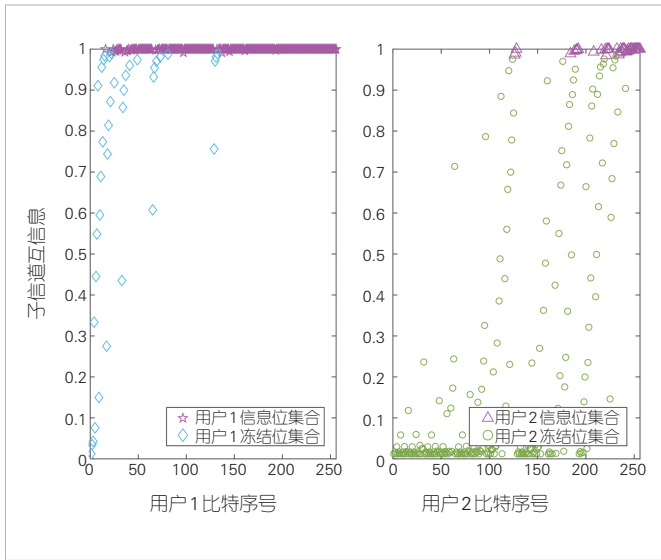
$$M_{\psi_1 \otimes \psi_2} = \frac{1}{2N} \sum_{k=1}^N (I_N^{(1,k)} + I_N^{(2,k)}) = M_{\psi_1},$$

$$V_{\psi_1 \otimes \psi_2} = \frac{1}{2N} \sum_{k=1}^N ((I_N^{(1,k)})^2 + (I_N^{(2,k)})^2) - M_{\psi_1 \otimes \psi_2}^2$$

$$\leq M_{\psi_1} (1 - M_{\psi_1}) \quad (8)$$

其中, $I_N^{(1,k)}$ 、 $I_N^{(2,k)}$ 分别是用户1和2的极化子信道互信息。可以看出,互信息均值在第二级变换保持不变,并且给出极化子信道互信息的方差上界,该上界由第一级功率分配决定。

对于极化编码系统来讲,极化子信道互信息方差能够在一定程度上反应编码系统的极化效应。级联系统的极化效应与功率分配密切相关,改善功率分配方案可以从极化的角度改善系统性能。所以功率域极化多址优化应该兼顾系统容量和极化效应。图5给出了每个用户码长 $N=256$,系统平均码率 $R=0.5$,用户信道系数 $|h_1|:|h_2|=3:1$,译码使用



▲图5 上行信道下功率分配导致的极化效应示意图

串行干扰抵消, 译码顺序为用户1-用户2, 情况下功率分配后极化效果示意图。

从图5中可以看出, 两用户的码率分配有明显差异。使用极化功率分配时, 可以轻松对所有子信道进行可靠性评估, 进而确定两用户的码率分配。通过对功率的分配和对应码率的计算, 获得更好的平均误码率性能。

3 无源多址接入技术

以保障用户链路可靠性为目标, 无源多址接入可以提高系统接入能力。本节首先介绍时隙擦除信道下的时隙极化变换 (SPT) 及其导致的时隙极化现象, 然后提出极化时隙ALOHA (PSA) 随机接入方法。

3.1 时隙极化变换及时隙极化现象

对于时隙擦除信道 (SEC) 而言, 在面向数据包操作 $\boxplus$ 下, 使用ARIKAN的极化内核 $\mathbf{F}_2$ 对独立的若干SEC信道副本进行变换可递归地生成包级的合并信道。即两个独立的SEC信道 S 经极化变换后得到第一级的合并信道 S_2 : $(\mathcal{X}^r)^2 \rightarrow (\mathcal{Y}^r)^2$, 其转移概率为:

$$S_2(\mathbf{y}_1, \mathbf{y}_2 | \mathbf{u}_1, \mathbf{u}_2) \triangleq S(\mathbf{y}_1 | \mathbf{u}_1 \boxplus \mathbf{u}_2) S(\mathbf{y}_2 | \mathbf{u}_2). \quad (9)$$

该合并信道 S_2 经拆分后得到的两个极化子信道 S^+ 和 S^- 分别为:

$$\begin{aligned} S^+(\mathbf{y}_1, \mathbf{y}_2, \mathbf{u}_1 | \mathbf{u}_2) &\triangleq \frac{1}{q} S_2(\mathbf{y}_1, \mathbf{y}_2 | \mathbf{u}_1, \mathbf{u}_2) \\ S^-(\mathbf{y}_1, \mathbf{y}_2 | \mathbf{u}_1) &\triangleq \sum_{\mathbf{u}_2 \in \mathcal{X}^r} \frac{1}{q} S_2(\mathbf{y}_1, \mathbf{y}_2 | \mathbf{u}_1, \mathbf{u}_2), \end{aligned} \quad (10)$$

其中, $\mathbf{u}_1, \mathbf{u}_2 \in \mathcal{X}^r, \mathbf{y}_1, \mathbf{y}_2 \in \mathcal{Y}^r$ 。

引理1: 基于数据包操作 $\boxplus$ 的合并信道 S_2 可以看作是由基于模-2 和操作的合并信道 W_2 的 r 个相同副本的形成的乘积信道, 即:

$$S_2(\mathbf{y}_1, \mathbf{y}_2 | \mathbf{u}_1, \mathbf{u}_2) = \prod_{i=1}^r W_2(\mathbf{y}_{1,i}, \mathbf{y}_{2,i} | \mathbf{u}_{1,i}, \mathbf{u}_{2,i}). \quad (11)$$

于是, 这两个极化子信道变为:

$$\begin{aligned} S^+(\mathbf{y}_1, \mathbf{y}_2, \mathbf{u}_1 | \mathbf{u}_2) &= \prod_{i=1}^r W^+(\mathbf{y}_{1,i}, \mathbf{y}_{2,i}, \mathbf{u}_{1,i} | \mathbf{u}_{2,i}) \\ S^-(\mathbf{y}_1, \mathbf{y}_2 | \mathbf{u}_1) &= \prod_{i=1}^r W^-(\mathbf{y}_{1,i}, \mathbf{y}_{2,i} | \mathbf{u}_{1,i}). \end{aligned} \quad (12)$$

从引理1可以看出: 包级的合并信道 S_2 等价于 r 个相同合并信道 W_2 所构成的乘型信道。同理, 我们对独立的两组上一级的合并信道继续递归地执行这种变换, 即可得到下一级合并信道: 两组第一级的极化子信道 S^+, S^- 再执行一次极化变换即可生成4个第二级的极化子信道 $S^{t_1 t_2}$, $t_1, t_2 \in \{+, -\}$ 。该极化变换递归的执行 n 次变换, 即可得到 $N = 2^n$ 个极化子信道 $S_N^{(j)}(\mathbf{y}_1^N, \mathbf{u}_1^{j-1} | \mathbf{u}_j)$, $j = 1, \dots, N$, 其表达式为:

$$S_N^{(j)} = \prod_{i=1}^r W_N^{(j)}(\mathbf{y}_{1,i}, \dots, \mathbf{y}_{N,i}, \mathbf{u}_{1,i}, \dots, \mathbf{u}_{j-1,i} | \mathbf{u}_{j,i}), \quad (13)$$

即第 n 级合并信道 S_n 等价于 r 个相同副本的合并信道 W_n 构成的乘型信道。

引理2: 时隙ALOHA随机接入系统时隙擦除信道信道传输时, $N = 2^n$ 个独立的SEC信道 $S(\epsilon)$ 经包级极化变换后得到的极化子信道的容量和Bhattacharyya参数分别为:

$$I(S_N^{(j)}) = rI(W_N^{(j)}) \quad (14)$$

和

$$Z(S_N^{(j)}) = rZ(W_N^{(j)}), \quad (15)$$

其中 $1 \leq j \leq N$ 。

基于该引理的结论, 可得SEC的合并信道容量和Bhattacharyya参数的递归计算关系式, 具体如下定理所示。

定理1: 对于 $1 \leq j \leq N/2$ 和 $N = 2^n$, SEC信道经SPT变换后的子信道容量可以通过下述递归计算求得:

$$\begin{cases} I(S_N^{(2^{j-1})}) = \frac{1}{r} I(S_{N/2}^{(j)})^2 \\ I(S_N^{(2^j)}) = 2I(S_{N/2}^{(j)}) - \frac{1}{r} I(S_{N/2}^{(j)})^2, \end{cases} \quad (16)$$

以及子信道Bhattacharyya参数的递归计算式:

$$\begin{cases} Z(S_N^{(2j-1)}) = 2Z(S_N^{(j)}) - \frac{1}{r}Z^2(S_N^{(j)}) \\ Z(S_N^{(2j)}) = \frac{1}{r}Z^2(S_N^{(j)}) \end{cases} \quad (17)$$

随着 N 的增加, SEC 信道经 SPT 变换后的子信道 $S_N^{(j)}$ 中一个部分子信道的对称容量趋向于 1, 而剩余的极化子信道则完全相反, 其信道的对称容量趋于 0。也就是说, 对于任何 $0 < \gamma < 1$,

$$\lim_{n \rightarrow \infty} \frac{|\{t \in \{+, -\}^n : I(S_N^{(t)}) \in (\gamma, 1 - \gamma)\}|}{2^n} = 0 \quad (18)$$

该定理的证明过程详见文献[9]。

由定理可以看出, 随着 N 的增加, 一些子信道 $S_N^{(i)}, i \in \mathcal{I}$ 的容量趋于 r (比特/信道), 而剩余的子信道容量趋向 0, 也就是说 SEC 的合并信道中的子信道 $S_N^{(i)}$ 也存在极化现象。在极化时隙随机接入中, 活跃用户是将集合 $\mathcal{I}$ 所对应的时隙子集作为接入方案随机选取的时隙集。

3.2 PSA 随机接入方法

本文提出的 PSA 随机接入系统如图 6 所示, 其中每个发送的数据包首先在各自的活跃用户端进行 PSA 接入的级联码编码; 当每个活跃用户在接收到基站 (BS) 广播的信标后, 将编码后的数据包经时隙擦除信道传输后在 BS 端构成一个传输帧。其中等效的 PSA 接入级联编码过程分为两个阶段, 可描述为:

$$\mathbf{x}_1^N = \mathbf{p}_1^M \mathbf{T} \mathbf{F}_2^{\otimes n} \quad (19)$$

对应地, 在 PSA 随机接入系统的接收 BS 端, 活跃用户发送过来数据包的译码恢复过程也分为两个阶段。

在 PSA 随机接入过程编码的第 1 阶段, 每个活跃用户 $i, 1 \leq i \leq M$, 在不规则度分布 $\Lambda(x)$ 指导下, 随机地从时隙子集 $\mathcal{I}$ 中为各自的数据包 $\mathbf{p}_i$ 的副本选择互异的时隙。然后生成时隙子集上的时隙数据包向量 $\mathbf{u}_T$:

$$\mathbf{u}_T = \mathbf{p}_1^M \cdot \mathbf{T} \quad (20)$$

生成矩阵 $\mathbf{T}$ 是从时隙 ALOHA 随机接入过程的生成矩阵集合 $\{\mathbf{T}\}$ 中随机选择的。

在第 2 阶段, 序号在 $\mathcal{I}^c$ 内时隙上的数据包被设置为 0, 从而得到等价的源数据包向量 $\mathbf{u}_1^N = (\mathbf{u}_T, \mathbf{u}_{T^c})$ 。接着基于包级的 SPT 变换, 生成时隙编码的数据包为 $\mathbf{x}_1^N$: $\mathbf{x}_1^N = \mathbf{u}_1^N \cdot \mathbf{F}_2^{\otimes n}$ 。

定义 PSA 随机接入方法中包级极化码码率为:

$$R_p \triangleq \frac{|\mathcal{I}|}{N} \quad (21)$$

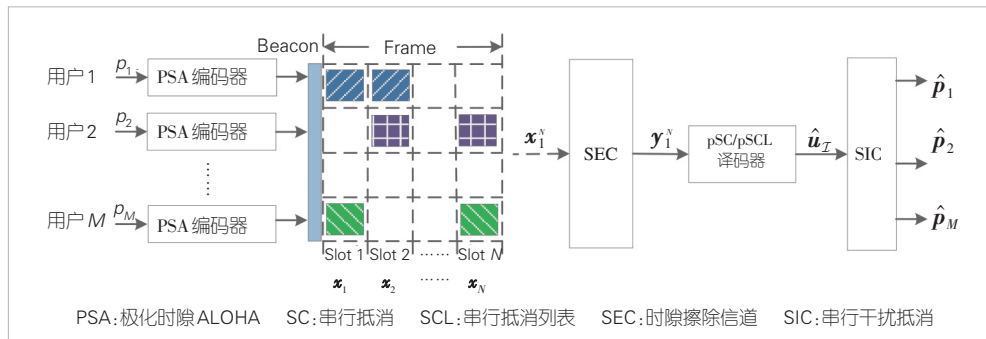
该参数是 PSA 随机接入方法中需优化的关键参数之一。

同理, 在 BS 端 PSA 随机接入方法的译码过程也分为两个阶段, 如图 3 中所述: 为了减少 SEC 造成的编码数据包的损坏, 首先将经信道传输后输出的向量 $\mathbf{y}_1^N$ 馈入到包级的 pSC/pSCL 译码器, 其译码后的数据包 $\hat{\mathbf{u}}_{\mathcal{I}}$ 经 SIC 过程迭代恢复发送的数据包 $\{\hat{\mathbf{p}}_i\}, 1 \leq i \leq M$, 详细的译码过程可参见文献[9]。

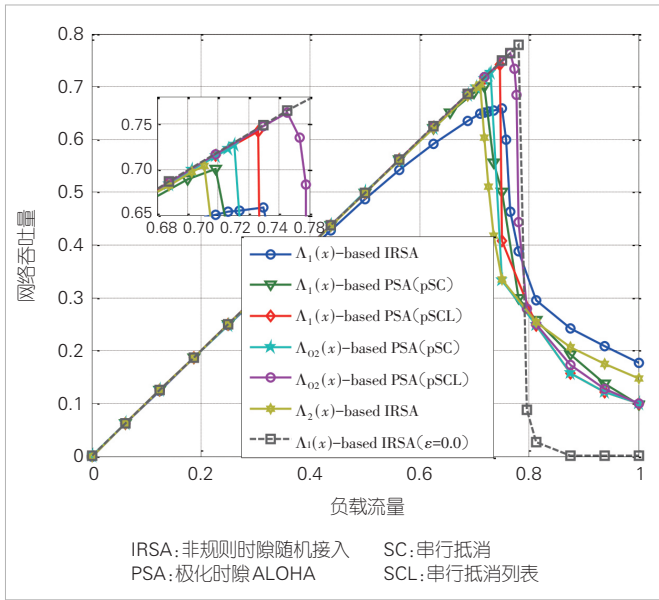
图 7 展示出了 SEC ($\epsilon = 0.1$) 信道下 PSA 和非规则时隙随机接入 (IRSA) 方案吞吐率对比性能曲线, 其中时隙帧长 $N = 1024$ 。图中还给出了度分布为 $\Lambda_1(x)$ 的 IRSA 随机接入方案在无擦除信道上传输的吞吐率基准性能曲线。仿真采用 $R = 1/3$ 的 3 个不同度分布, 分别选取 $\Lambda_1(x) = 0.554016x^2 + 0.261312x^3 + 0.184672x^6$ (对应的阈值为 $G_1^* = 0.8792$), 针对 SEC ($\epsilon = 0.1$) 信道优化的度分布 $\Lambda_2(x) = 0.0695x^2 + 0.8958x^3 + 0.0347x^5$ (对应的阈值为 G_2^*), 以及本文优化的度分布 $\Lambda_{02}(x) = 0.5610x^2 + 0.2002x^3 + 0.1554x^5 + 0.0029x^6$ (对应的阈值为 $G_3^* = 0.8898$) 用于 PSA 随机接入系统性能的评估。

从图中可以看出: 1) 在同一用户节点 (UN) 度分布 $\Lambda_1(x)$ 下, 与无擦除信道上传输时 IRSA 的峰值吞吐率相比, IRSA 随机接入方案经 SEC 信道传输造成的吞吐率峰值的损失约为 0.14。同样在前述相同的度分布 $\Lambda_1(x)$ 下, 经 SEC 信道传输的 PSA 随机接入系统

(采用 pSCL($L = 8$) 译码算法) 吞吐率峰值的提高量约为 0.1。2) 在采用本文优化的 UN 度分布 $\Lambda_{02}(x)$ 下, 使用 pSCL($L = 8$) 译码的 PSA 随机接入方案的吞吐率峰值提升了约 0.12, 仅比基于 $\Lambda_1(x)$ 的 IRSA 随机接入方案在无差错信道下的基准峰值吞吐率低 0.02。



▲图 6 经 SEC 信道传输的 PSA 随机接入系统模型



▲图7 PSA随机接入方案经时隙擦除信道传输的系统吞吐量随输入负载变化的性能对比图

除上述方案外,文献[11]提出的基于极化编码的高斯多址接入信道(GMAC)信道无源多址接入方案,使用联合检测和译码算法,能够有效贴近编码时隙ALOHA(CSA)的理论界。文献[12]使用随机扩频的方案来减少多用户间干扰,进一步提高了极化随机接入系统的接入能力。通过上述描述,可以看出,无源极化多址接入系统是一种提高6G系统吞吐率的有效手段,是有未来6G无线接入方案有力的候选方案。

4 结束语

本文提出了多种极化多址接入系统,包括有源多址接入和无源多址接入。有源极化多址接入包括编码域NOMA和功率域NOMA方案,利用系统的极化效应来提升系统容量。无源极化多址方案,以编码随机接入为基础,证明了SEC经SPT变换后的合成子信道中存在极化现象,本文所提出的PSA随机接入方案可以有效提升系统的接入能力。

综上所述,基于极化编码的多址接入技术,是满足未来6G移动通信需求的重要候选技术,具有广阔的应用前景。

参考文献

- [1] ITU-R. Framework and overall objectives of the future development of IMT for 2030 and beyond [R]. 2023
- [2] POLYANSKIY Y. A perspective on massive random-access [C]// Proceedings of IEEE International Symposium on Information Theory (ISIT). IEEE, 2017: 2523–2527. DOI: 10.1109/isit.2017.8006984
- [3] ARIKAN E. Channel polarization: a method for constructing

capacity-achieving codes for symmetric binary-input memoryless channels [J]. IEEE transactions on information theory, 2009, 55(7): 3051–3073. DOI: 10.1109/TIT.2009.2021379

- [4] 牛凯. 极化码原理与应用 [M]. 北京: 科学出版社, 2021
- [5] NIU K, CHEN K. CRC-aided decoding of polar codes [J]. IEEE communications letters, 2012, 16(10): 1668–1671. DOI: 10.1109/lcomm.2012.090312.121501
- [6] LI B, SHEN H, TSE D. An adaptive successive cancellation list decoder for polar codes with cyclic redundancy check [J]. IEEE communications letters, 2012, 16(12): 2044–2047. DOI: 10.1109/lcomm.2012.111612.121898
- [7] PIAO J N, DAI J C, NIU K. CRC-aided sphere decoding for short polar codes [J]. IEEE communications letters, 2019, 23(2): 210–213. DOI: 10.1109/lcomm.2018.2885771
- [8] DAI J C, NIU K, SI Z, et al. Polar-coded non-orthogonal multiple access [J]. IEEE transactions on signal processing: a publication of the IEEE signal processing society, 2017, 66(5): 1374–1389. DOI: 10.1109/TSP.2017.2786273
- [9] ZHANG Z J, NIU K, DAI J C, et al. Polar-slotted ALOHA over slot erasure channel [J]. IEEE transactions on vehicular technology, 2023, 72(1): 760–771. DOI: 10.1109/tvt.2022.3204321
- [10] SUN Z, XIE Y X, YUAN J H, et al. Coded slotted ALOHA for erasure channels: design and throughput analysis [J]. IEEE transactions on communications, 2017, 65(11): 4817–4830. DOI: 10.1109/tcomm.2017.2734649
- [11] MARSHAKOV E, BALITSKIY G, ANDREEV K, et al. A polar code based unsourced random access for the Gaussian MAC [C]// Proceedings of IEEE 90th Vehicular Technology Conference (VTC2019-Fall). IEEE, 2019: 1–5. DOI: 10.1109/vtcfall.2019.8891583
- [12] PRADHAN A K, AMALLADINNE V K, NARAYANAN K R, et al. Polar coding and random spreading for unsourced multiple access [C]// Proceedings of ICC 2020 – 2020 IEEE International Conference on Communications (ICC). IEEE, 2020: 1–6. DOI: 10.1109/icc40277.2020.9148687

作者简介



崔宏基, 北京邮电大学在读博士研究生; 主要研究方向为5G/6G移动通信、信道编码和多址接入。



牛凯, 北京邮电大学人工智能学院教授、博士生导师, 北京邮电大学泛网无线通信教育部重点实验室副主任, 中国电子学会信息论分会副主任委员; 主要研究方向为5G/6G移动通信、智能信号处理; 先后主持多项国家自然科学基金重点与面上项目、重点研发计划项目; 荣获中国电子学会科技奖自然科学一等奖; 发表论文200余篇, 申请国家发明专利100余项, 所提极化码高可靠编码方案被写入5G标准。

无蜂窝大规模MIMO中的大规模随机接入



Massive Random Access in Cell-Free Massive MIMO System

胡彦丰/HU Yanfeng¹, 王东明/WANG Dongming¹,
梁楚龙/LIANG Chulong², 尤肖虎/YOU Xiaohu¹

(1. 东南大学, 中国 南京 211102;
2. 中兴通讯股份有限公司, 中国 深圳 518057)

(1. Southeast University, Nanjing 211102, China;
2. ZTE Corporation, Shenzhen 518057, China)

DOI: 10.12142/ZTETJ.202401006

网络出版地址: <http://kns.cnki.net/kcms/detail/34.1228.TN.20240226.0948.004.html>

网络出版日期: 2024-02-26

收稿日期: 2024-01-16

摘要: 研究了以用户为中心的无蜂窝大规模多输入多输出 (MIMO) 架构下的大规模随机接入方案。为实现可扩展架构, 讨论了接入点 (AP) 与用户设备 (UE) 的关联以及 AP 的分簇方法。针对活跃用户检测 (AUD), 一种基于最大似然检测 (ML) 的方案被提出以获取活跃用户。通过调整阈值, 可以得到不同精度的检测结果。利用 AUD 的检测用户集合, 系统利用基于狄利克雷过程的稀疏贝叶斯学习 (DP-SBL) 完成信道估计 (CE)。该算法可以充分利用 AP 的空间聚散特性, 提高系统准确性。基于以上工作, 我们提出了联合 AUD 和 CE 算法。仿真结果验证了所提方案在性能上的优越性。

关键词: 大规模随机接入; 无蜂窝大规模 MIMO; 活跃用户检测; 信道估计

Abstract: A user-centric massive random access scheme under the context of cell-free massive multiple-input multiple-output (MIMO) architecture is investigated. To achieve a scalable architecture, the association between access points (APs) and user equipment (UE) as well as the clustering method for APs is discussed. Regarding active UE detection (AUD), a class of maximum likelihood (ML)-based schemes is proposed to obtain active UE set. By adjusting the threshold, detection results with varying accuracies can be achieved. Leveraging the detected user set from AUD, the system employs sparse Bayesian learning based on Dirichlet process (DP-SBL) for channel estimation (CE), effectively utilizing the spatial clustering characteristics of APs to enhance accuracy. Building on this, a joint AUD and CE algorithm is proposed. Simulation results validate the superiority of the proposed approach in terms of performance.

Keywords: massive random access; cell-free massive MIMO; active UE detection; channel estimation

引用格式: 胡彦丰, 王东明, 梁楚龙, 等. 无蜂窝大规模MIMO中的大规模随机接入 [J]. 中兴通讯技术, 2024, 30(1): 26-32. DOI: 10.12142/ZTETJ.202401006

Citation: HU Y F, WANG D M, LIANG C L, et al. Massive random access in cell-free massive MIMO system [J]. ZTE technology journal, 2024, 30(1): 26-32. DOI: 10.12142/ZTETJ.202401006

大规模机器通信 (mMTC) 是 5G 移动通信技术的关键应用场景之一。在网络中支持海量 MTC 设备 (MTCD) 的大规模随机接入情景下, 如何有效地估计信道、检测活跃用户和对应的数据消息逐渐成为热点研究课题。在 mMTC 场景中, 系统可接入的用户设备量极大, 这时由于可用信道资源是有限的, 以往基于正交资源分配的方案将限制设备的接入效率。在这种背景下, 多个用户利用特殊的编码调制方案或特定的签名, 接收机可由对应的算法分离多用户信息。由于这里使用了非正交时频资源, 所以这种方式也可以称为非

正交多址接入 (NOMA) [1-3]。现有的随机接入方案中, 设备需要通过物理随机接入信道 (PRACH) 和基于争用的四路握手过程实现接入。LTE 系统四路握手过程包括: 前导信号传输、随机接入响应、无线电资源控制连接请求、建立无线电资源控制 (RRC) 请求 [4]。该过程又称为资源授权过程, 是影响 mMTC 效率的主要原因 [5]。因此, 基于无授权的大规模接入将会是未来 mMTC 场景的关键技术方案之一。

近年来, 为了消除小区边际效应, 进一步提高通信系统的传输效率, 文献 [6] 提出了无蜂窝大规模多输入多输出 (MIMO) 系统。E. BJORNSSON 等认为无蜂窝大规模 MIMO 是网络 MIMO 架构下动态协作分簇 (DCC) 的一个特例 [7]。由

基金项目: 国家重点研发计划 (2020YFB1807200)

此可以根据DCC的方案,开发用于联合初始接入、导频分配和协作集群形成的新的以用户为中心的可扩展算法^[8]。基于无蜂窝大规模MIMO架构,M. L. KE等提出了对应的大规模随机接入方案^[9-10]。这些方案均假设用户与所有AP相关联,在超密集设备分布情况下不具备可扩展性,并且假设AP架设高度远高于UE^[11],以引入经典的单环信道模型来形成角度域的稀疏特性^[12]。然而通常情况下,该假设并不完全适用。因此,本文将进一步研究以用户为中心的无蜂窝大规模MIMO随机接入方案。

1 系统模型

无蜂窝网络的关键特征是可扩展性,即协作接入点和用户设备(UE)的维度可以扩展到无穷大。在传统的可扩展无蜂窝大规模MIMO中,每个AP通过利用本地信道状态信息(CSI)来处理配对UE的数据。通常AP会选择关联的用户,而每个AP的关联容量有限,这样在超密集部署MTCD的场景下,就会存在遗漏用户未被关联的情况。此外,全分布式的无蜂窝大规模MIMO存在前传开销大、协作能力有限等问题。灵活的协同信号处理架构不仅可以实现较好的传输性能,同时也具备可扩展性。在本文中,我们将考虑以用户为中心的可扩展性无蜂窝大规模MIMO架构,即在无蜂窝无线接入网(RAN)架构中,增加边缘分布式单元(EDU)模块,并使之与多个AP连接,从而增强联合处理能力,并在EDU上进行分布式接收/预编码,实现无蜂窝大规模MIMO网络的无限扩展。此外,以用户为中心的分布式单元(UCDU)可实现用户数据的分布式协作合并和分发。如图1所示,新型全频谱无蜂窝无线接入网架构主要包括AP、EDU、UCDU和CPU,其相应特点为^[13]:

- 1) AP完成射频信号的收发、数模/模数转换等功能;
- 2) EDU主要完成基带信号处理中的信道估计、多用户/多数据流检测、多用户/多数据流预编码、校准信号的提取等功能;
- 3) UCDU主要实现将收到的多个EDU发送来的上行数据进行合并,将多个下行数据流分发到相应的EDU,AP之间校准所需的信道估计等功能;
- 4) CPU实现确定EDU和UCDU的数据流对应关系等功能。

在本文中,我们主要考虑上行mMTC场景中的活跃用户检测与信道估计,因此涉及的信号处理模块只包括AP和EDU。用户与AP关联以及EDU与AP的关联方法是本架构信号处理过程的重要实现步骤,相关的内容将在本文后续章节中完成进一步阐述。

2 设备预关联

在接入区域中,系统将提供有限数量的计算节点(即EDU),但与之相对应的有较大数量的AP。因此,通过AP分簇可以让少量的EDU处理较多AP的接收信号。Kmeans++算法被认为是一种行之有效的分簇方法。在给定EDU数量后,系统根据AP的位置,使得各AP簇中心之间距离和达到最大以便完成分簇。

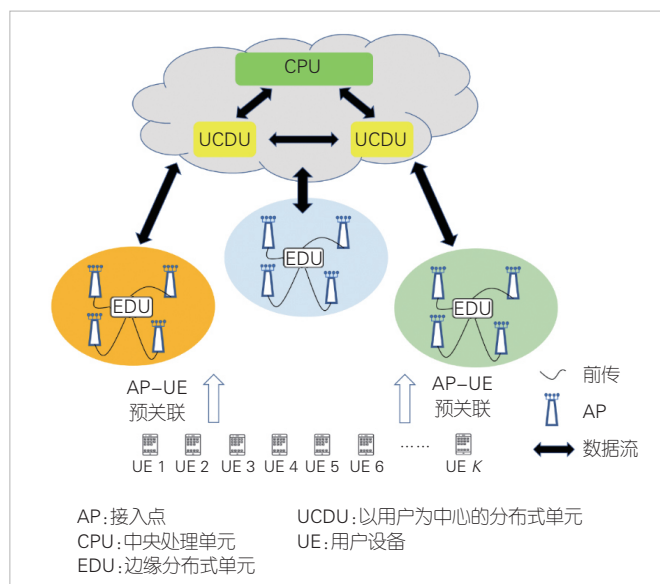
类似于DCC的初始接入过程^[14],系统以UE为中心,让UE选择AP以完成关联。这样UE可以通过更好的信道传输信息。AP根据自身被关联的情况选择是否完成与该UE的关联。最后的关联关系将发送给EDU。

设备预关联阶段发生在设备完成接入之前。通常我们假设UE的位置保持相对不变。此外,考虑到计算资源和存储资源的稀缺性,以及系统应当具备的可扩展性,由AP聚类完成部分联合信号处理的策略是可行的。因此,在UE随机接入之前,各AP应当已经完成分簇,并且UE已完成与AP的关联,以此为基础实现下一阶段处理。本文所探究的重点不在预关联,因此不着重展开介绍这部分内容,并假设设备预关联已经完成。

3 最大似然活跃检测

本文所考虑的信道为瑞利衰落信道,第**b**个AP的接收前导信号矩阵 $\mathbf{Y}_b$ 的每一列都是独立同分布的的向量,其均值和方差分别表示为:

$$E\left\{\left[\mathbf{Y}_b\right]_{:,m}\right\}=\mathbf{0}, \quad (1)$$



▲图1 新型无蜂窝接入网架构模型图

$$V\{[Y_b]_{:,m}\} = \sum_k \theta_{k,b} \Phi_k \Phi_k^H + \sigma^2 I_L, \quad (2)$$

其中 $\theta_{k,b} = \alpha_k \lambda_{k,b}$, 表示用户 k 到第 b 个 AP 的活跃信道大尺度系数。以上结果对任意 $m \in \{1, \dots, M\}$ 都成立, M 为 AP 天线数量。则我们令 $\Xi_b = V\{[Y_b]_{:,m}\}$, 满足 $[Y_b]_{:,m} \sim \mathcal{CN}(\mathbf{0}, \Xi_b)$ 。令 $\theta_b = [\theta_{1,b}, \dots, \theta_{K,b}]^T$, 利用信道向量与接收信号向量的高斯特性, 我们将对 θ_b 进行最大似然估计。首先我们列出 θ_b 的对数似然损失函数:

$$c(\theta_b) = -\frac{1}{M} \log p(Y_b | \theta_b) = -\frac{1}{M} \sum_{m=1}^M \log p([Y_b]_{:,m} | \theta_b) = \log |\Xi_b| + \text{tr}(\Xi_b^{-1} \hat{\Xi}_b), \quad (3)$$

其中,

$$\hat{\Xi}_b = \frac{1}{M} Y_b Y_b^H = \frac{1}{M} \sum_{m=1}^M [Y_b]_{:,m} [Y_b]_{:,m}^H. \quad (4)$$

该矩阵为接收信号矩阵 $\mathbf{Y}$ 的采样协方差矩阵。当 M 趋向于无穷大的时候, $\hat{\Xi}_b$ 趋向于 Ξ_b 。根据最大似然估计, 理想情况下活跃用户检测结果应该表示为:

$$\theta_{b,ML} = \underset{\{\theta_{k,b}\} \text{ for } k \in \{K_n\}}{\arg \min} c(\theta_b). \quad (5)$$

n 为当前 AP 所在的 EDU 索引, $\{K_n\}$ 表示该第 n 个 EDU 覆盖下的 AP 关联 UE 集合, $\theta_b \in \mathbb{R}_+^{K_n \times 1}$ 。于是, 最大似然多用户检测过程被总结为算法 1。

算法 1: 活跃用户检测算法

输入: 第 b 个 AP 的接收信号矩阵 $\mathbf{Y}_b$, 第 n 个 EDU 簇内关联 UE 的 Preamble 序列 Ψ_n , 噪声方差 σ^2 , 精确活跃用户检测阈值 th_a , 粗略活跃用户检测阈值 th_r , 迭代次数 I 。

输出: 精确活跃用户集合 $\{S_a: k | \theta_{k,b} > th_a\}$, 粗略活跃用户集合 $\{S_r: k | \theta_{k,b} > th_r\}$ 。

1: 初始化计算: $(\tilde{\Xi}_b^{-1})^1 = \frac{1}{\sigma^2} I_L, (\theta_b)^1 = \mathbf{0}, \hat{\Xi}_b = \frac{1}{M} Y_b Y_b^H, i=1$;

2: 当 $i < I$ 时, 执行以下操作;

3: 随机选取索引 $k' \in \{K_n\}$, 对应 θ_b 的第 k' 个元素;

4: 计算 $t^* =$

$$\max \left[\frac{\Phi_k^H(\tilde{\Xi}_b^{-1})^{i-1} \hat{\Xi}_b (\tilde{\Xi}_b^{-1})^{i-1} \Phi_{k'} - \Phi_{k'}^H(\tilde{\Xi}_b^{-1})^{i-1} \Phi_{k'}}{(\Phi_k^H(\tilde{\Xi}_b^{-1})^{i-1} \Phi_{k'})^2}, -\theta_{k',b} \right];$$

5: 更新 $(\theta_b)^{i+1} = (\theta_b)^i + t^* \cdot e_{k'}$;

$$6: \text{更新 } (\tilde{\Xi}_b^{-1})^i = (\tilde{\Xi}_b^{-1})^{i-1} - \frac{t^* (\tilde{\Xi}_b^{-1})^{i-1} \Phi_{k'} \Phi_{k'}^H (\tilde{\Xi}_b^{-1})^{i-1}}{1 + t^* \Phi_{k'}^H (\tilde{\Xi}_b^{-1})^{i-1} \Phi_{k'}};$$

7: 更新 $i = i + 1$;

8: 结束

4 基于 DP-SBL 的信道估计方法

在本节中, 我们将对信道估计方法进行探究。在未知信道的先验信息的情况下, 我们通过稀疏贝叶斯学习 (SBL) 的方法对信道超参数信息进行估计, 并利用 AP 的空间聚类特性完成对这些超参数的分簇。如图 2 所示, 空间位置相近的 AP 信道向量具有相类似的稀疏模式, 因此 AP1 和 AP2 的超参数可以划分到同一个簇当中。

狄利克雷过程被认为是“概率分布的分布”, 由基础分布 G_0 和正数放缩参数 η 确定, 因此可以被写成 $G = DP(\eta, G_0)$ 。本文采用断棍结构, 将 G 的表达式以狄拉克函数的加权的方式进行表达。 G 具体表示为:

$$G = \sum_{x=1}^{\infty} \omega_x \delta(\hat{\gamma}^{(x)}), \quad (6)$$

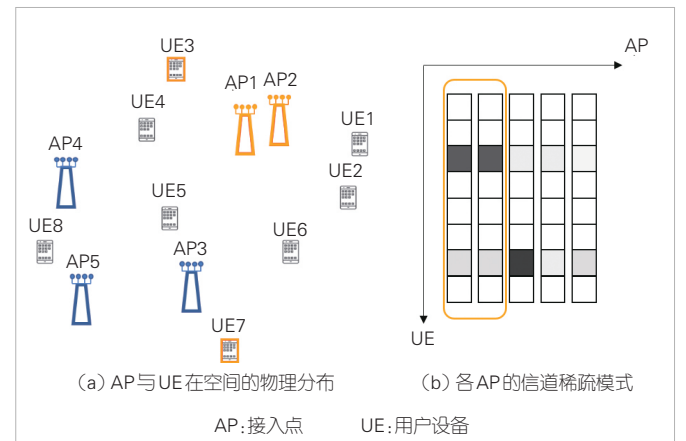
其中,

$$\omega_x = \pi_x \prod_{i=1}^{x-1} (1 - \pi_i), \quad (7)$$

$$p(\pi_x | \eta) = \text{Be}(\pi_x; 1, \eta), \quad (8)$$

$$p(\hat{\gamma}^{(x)}) = G_0. \quad (9)$$

这里的 $\{\hat{\gamma}^{(x)}, x=1:\infty\}$ 是根据分布 G_0 获得的采样向量集合, $\text{Be}(x; a, b)$ 表示 x 服从参数为 a, b 的贝塔分布。由于实际的处理中不可能获取无限的样本, 因此我们将样本数量设



▲图 2 空间距离相近的 AP 具有相似的信道稀疏模式

置为足够大的值 X (X 小于总天线数量)。在此基础上,我们将SBL算法中用户信道方差的倒数的先验分布由伽玛分布替换为狄利克雷过程,即可得到DP-SBL的信道估计算法。

5 EDU的信号处理过程

5.1 单次联合策略

活跃用户检测(AUD)和信道估计(CE)的计算过程都在EDU处完成。在系统完成设备预关联之后,EDU将获取簇内AP所关联UE的标签信息,包括UE索引和对应的前导序列。EDU先对每个AP进行活跃用户检测,分别得到粗略活跃用户集合和精确活跃用户集合。设置合适的阈值,使得各AP精确活跃用户集合合并为AUD结果。将各AP粗略用户集合的并集输入到CE模块,以此获取CE的测量矩阵,完成基于狄利克雷过程(DP)的SBL信道估计,最终输出各活跃用户信道估计结果。整个处理过程被总结为算法2。

算法2:EDU联合信号处理策略

输入:接收信号 $\{Y_b; b \text{ in cluster } n\}$ 。

输出:活跃用户检测集合 $\{S_a\}$ 和估计信道向量 $\{\bar{g}_a\}$ 。

- 1: 利用算法1得到各AP的精确活跃用户 $\{S_{b,a}\}$ 和粗略活跃用户集合 $\{S_{b,r}\}$;
- 2: 将各AP活跃用户集合取并集得到 $\{S_a = \bigcup_b S_{b,a}\}$ 和 $\{S_r = \bigcup_b S_{b,r}\}$ 。输出 $\{S_a\}$ 为活跃用户检测集合,将 $\{S_r\}$ 传递到CE模块;
- 3: 通过 $\{S_r\}$ 和关联用户标签,CE模块获取压缩感知测量矩阵 Ψ_n ;
- 4: 将 Ψ_n 和接收信号 $\{Y_b\}$ 输入到DP-SBL算法中,得到 $\{S_r\}$ 的估计信道向量 $\{\bar{g}_r\}$ 。由于 $\{S_a\} \subseteq \{S_r\}$,我们将精确活跃用户集合的估计信道向量 $\{\bar{g}_a\}$ 输出,从而完成CE。

5.2 基于SIC的联合策略

与4.1所提出的联合策略不同,EDU加入了串行干扰消除(SIC)模块,在每次SIC迭代的过程中减去精确活跃用户集的接收信号,从而不断降低信道向量的稀疏度,如图3所示。相比于单次处理,该策略的精确性会有所提高。除此之外,在每次SIC迭代中,灵活调整AUD精确阈值 th_a 和粗略阈值 th_r ,可使得压缩感知计算结果更为精准。通过加入SIC模块,系统的性能将会得到提升。然而,算法的精细化设计将带来较高的计算复杂度。

算法3:EDU基于SIC的联合信号处理策略

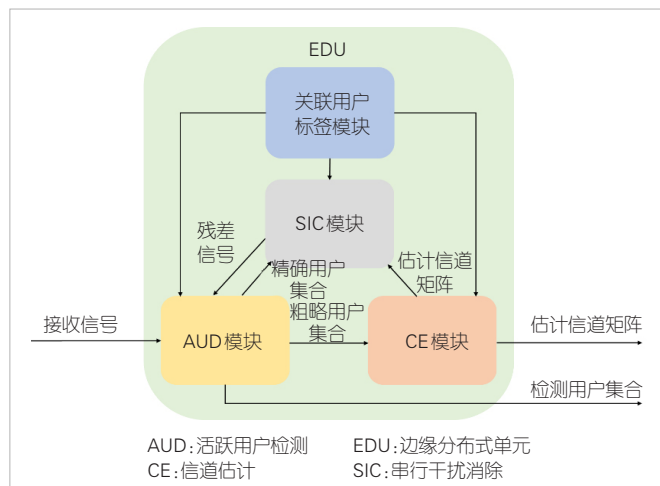
输入:输入接收信号 $\{Y_b\}$,其中 b 在第 n 个聚类内,SIC迭代次数 I_{SIC} ,初始化 $i = 1$ 。

输出:每次SIC迭代AUD模块得到的精确活跃用户集合 $\{S_a\}$ 和每次CE模块得到的精确活跃用户信道向量集合 $\{\bar{g}_a\}$ 。

- 1: 当 $i < I_{SIC}$ 时,执行以下操作;
- 2: 利用算法1得到各AP的精确活跃用户 $\{S_{b,a}\}$ 和粗略活跃用户集合 $\{S_{b,r}\}$;
- 3: 将各AP活跃用户集合取并集得到 $\{S_a = \bigcup_b S_{b,a}\}$ 和 $\{S_r = \bigcup_b S_{b,r}\}$ 。输出 $\{S_a\}$ 为活跃用户检测集合,并将 $\{S_a\}$ 和 $\{S_r\}$ 作为输入分别传递给SIC模块和CE模块;
- 4: 通过 $\{S_r\}$ 和关联用户标签,CE模块获取压缩感知测量矩阵 Ψ_n 。将 Ψ_n 和接收信号 $\{Y_b\}$ 输入到DP-SBL算法中,得到 $\{S_r\}$ 的估计信道向量 $\{\bar{g}_r\}$ 。将精确活跃用户集合的估计信道向量 $\{\bar{g}_a\}$ 输出,并作为输入传递给SIC模块;
- 5: SIC模块计算残差信号 $\{Y_b = Y_b - \Psi_{n,a} \cdot \bar{G}_a\}$, $\Psi_{n,a}$ 表示索引为 $\{S_a\}$ 的前导序列矩阵, $\bar{G}_a$ 表示由 $\{\bar{g}_a\}$ 排列形成的矩阵。SIC模块将残差信号传递给AUD模块;
- 6: 重新设置 $th_a = \alpha \cdot th_a$, $th_r = \beta \cdot th_r$, $i = i + 1$, 其中 $\alpha < 1$, $\beta < 1$, 并且均为常数;
- 7: 结束

6 数值仿真

本节我们将设置一些仿真来验证本文所提大规模随机接入方案的优越性。在无蜂窝场景中,UE与AP随机地分布在



▲图3 基于SIC的联合AUD和CE处理模块示意图

接入区域中。为了简化计算处理过程，我们假设存在单天线UE，并且上行发射功率都一致。每个用户的Preamble序列均为服从i.i.d的高斯向量。部分仿真参数可见于表1。

此外，AUD的检测错误率（DER）和CE的归一化平均误差（NMSE）用来衡量算法性能。它们分别被定义为：

$$DER = \frac{\sum_k |\hat{\alpha}_k - \alpha_k|}{K}, \quad (10)$$

$$NMSE = 10 \log_{10} \frac{\sum_{k^*} \sum_{b^*} \|\bar{\mathbf{g}}_{b^*, k^*} - \mathbf{g}_{b^*, k^*}\|_2^2}{\sum_{b^*} \sum_{k^*} \|\mathbf{g}_{b^*, k^*}\|_2^2}, \quad (11)$$

其中， $\hat{\alpha}_k$ 表示对第 k 个UE的活跃检测结论。若 $k \in \bigcup \{S_a\}$ ，则表明 $\hat{\alpha}_k = 1$ ，否则 $\hat{\alpha}_k = 0$ 。 k^* 为估计活跃用户集合与真实活跃用户集合的并集UE索引，即：

$$k^* \in \{\bigcup \{S_a\}\} \cup \{\bigcup \{S\}\}, \quad (12)$$

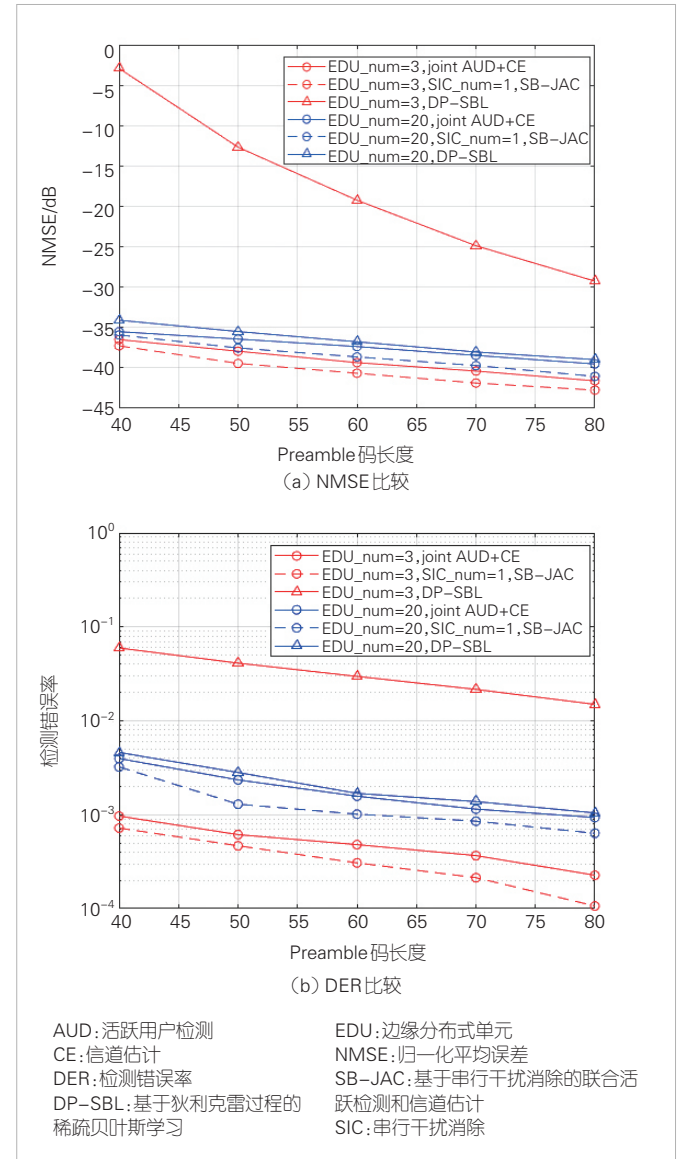
同时， b^* 为当前第 k^* 个UE所关联的AP索引。

图4比较了单次联合策略和基于SIC联合策略的信号处理性能。此外，我们将单独使用阈值DP-SBL完成AUD和CE的方法作为对照。从图中可以看出，在EDU数量较少的情况下，信号处理接近于全集中式处理，无论基于串行干扰消除的联合活跃检测和信道估计（SB-JAC）算法还是单次联合活跃检测和信道估计（JAC）算法的性能都优于EDU数

▼表1 仿真参数

参数	数值
接入区域半径	2.3 km
背景噪声谱密度	-174 dBm/Hz
系统带宽	10 MHz
UE数量	5 000
活跃UE数量	300
每个AP关联UE的数量	330
每个UE关联的AP数量	5
UE发送功率	18 dBm
AUD迭代次数 l_{AUD}	50
CE迭代次数 T	50
SIC迭代次数 l_{SIC}	3
联合策略的精确阈值 th_a	1.6
联合策略的粗略阈值 th_r	0.5
SIC联合策略的初始精确阈值 th_a	150
SIC联合策略的初始粗略阈值 th_r	1.0
大尺度系数 $\rho_{b,k}$ 与距离 $d_{b,k}$ 的关系	$128.1 + 37.6 \log_{10}(d_{b,k})$

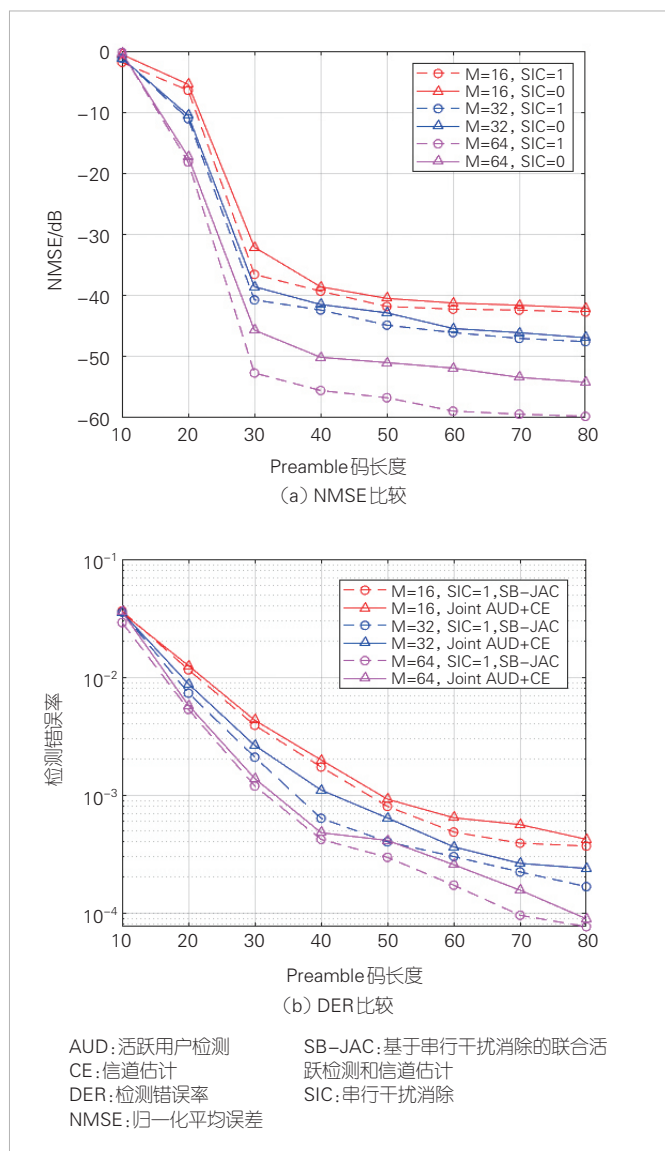
AP:接入点
AUD:活跃用户检测
CE:信道估计
SIC:串行干扰消除
UE:用户设备



▲图4 基于SIC的联合策略、单次联合策略和DP-SBL算法性能比较

量较多的情形。但是由于EDU所关联的用户数量增多，在前导序列长度有限的情况下，EDU_num=3时，单独DP-SBL算法的信号处理性能远不如其他两种算法。而在EDU_num=20时，信号处理近似于全分布式处理，JAC的协作处理能力不足，因此单独DP-SBL算法和JAC算法的性能较为接近。由于在SIC迭代过程中估计信道向量的稀疏度降低了，因此SB-JAC算法的性能要优于JAC算法，但是这同样增加了计算复杂度。

图5展示了天线数量对联合策略性能的影响。EDU数量为7，每个AP的天线数量 M 分别为16、32和64。在天线数量增多的情况下，基于最大似然估计的用户检测准确性得到提升，并且天线聚类的效应得到增强，因此随机接入的



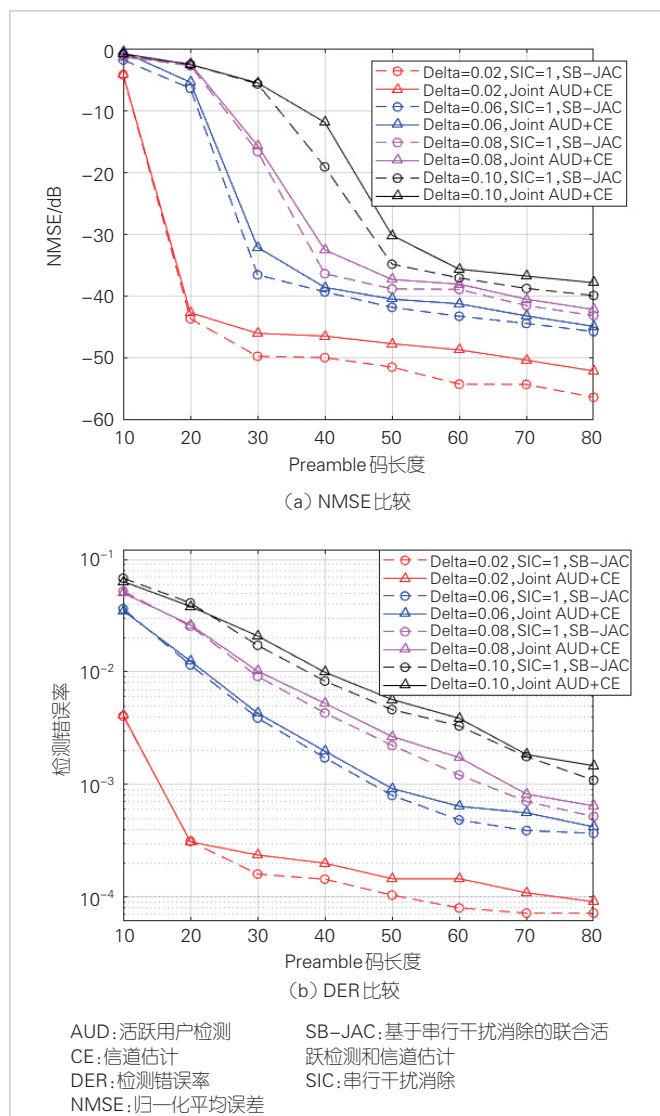
▲图5 基于SIC的联合策略、单次联合策略在AP天线数量变化时性能比较

AUD和CE性能都得到了改善,并且SIC的策略性能要优于单次联合性能。

图6比较了活跃用户密度对性能的影响,其中参数delta表示活跃用户数量占总体用户数量的比值。设置每个AP的天线数量为16,保持总体用户数量不变,我们可以看出活跃用户越稀疏,系统的性能表现就越好。该结论验证了压缩感知在低稀疏度下拥有更好的估计性能。仿真结果表明,SIC的策略性能优于单次联合性能。

7 结束语

基于以用户为中心的新型无蜂窝大规模MIMO架构,本文提出了一种大规模随机接入方案。在此架构下,我们首先



▲图6 基于SIC的联合策略、单次联合策略在用户稀疏度变化时性能比较

探讨了设备预关联方案,包括AP聚类和UE-AP关联。在此基础上,我们研究了基于最大似然检测的多用户活跃性检测算法和基于DP-SBL的信道估计算法,其中AUD算法不依赖信道的先验信息,且不进行矩阵求逆,可以在检测准确性和计算复杂度取得较好的折中。基于DP-SBL的CE算法可以充分利用AP分布的空间特性,适用于本文所研究的大规模分布式天线。结合以上算法,我们提出了联合AUD和CE的大规模随机接入信号处理策略。仿真表明,相比于非联合的DP-SBL算法,联合算法能够降低信道估计NMSE和活跃用户检测DER,特别是EDU数量较少的情况。在此基础上,我们提出了基于SIC的联合AUD和CE算法,仿真显示该算法可以进一步提高AUD和CE的性能,但同时会增加系统的计算复杂度。

在未来无线通信中, MTC将占据通信流量的七成以上。与无蜂窝技术的结合将会成为mMTC的主流研究方向。对于大规模随机接入方案的探究仍需要持续关注。

参考文献

- [1] DAI L L, WANG B C, YUAN Y F, et al. Non-orthogonal multiple access for 5G: solutions, challenges, opportunities, and future research trends [J]. IEEE communications magazine, 2015, 53(9): 74–81. DOI: 10.1109/MCOM.2015.7263349<http://dx.doi.org/10.1109/MCOM.2015.7263349>
- [2] TAO Y Z, LIU L, LIU S, et al. A survey: several technologies of non-orthogonal transmission for 5G [J]. China communications, 2015, 12(10): 1–15. DOI: 10.1109/CC.2015.7315054<http://dx.doi.org/10.1109/CC.2015.7315054>
- [3] WANG Y M, REN B, SUN S H, et al. Analysis of non-orthogonal multiple access for 5G [J]. China communications, 2016, 13(2): 52–66. DOI: 10.1109/CC.2016.7405722<http://dx.doi.org/10.1109/CC.2016.7405722>
- [4] DAHLMAN E, PARKVALL S, SKOLD J, et al. 3G evolution: HSPA and LTE for mobile broadband [M]. New York: Academic Press, 2008
- [5] CHEN H, ABBAS R, CHENG P, et al. Ultra-reliable low latency cellular networks: use cases, challenges and approaches [J]. IEEE communications magazine, 2018, 56(12): 119–125. DOI: 10.1109/MCOM.2018.1701178<http://dx.doi.org/10.1109/MCOM.2018.1701178>
- [6] NGO H Q, ASHIKHMIN A, YANG H, et al. Cell-free massive MIMO versus small cells [J]. IEEE transactions on wireless communications, 2017, 16(3): 1834–1850. DOI: 10.1109/TWC.2017.2655515<http://dx.doi.org/10.1109/TWC.2017.2655515>
- [7] BJÖRNSON E, SANGUINETTI L. Making cell-free massive MIMO competitive with MMSE processing and centralized implementation [J]. IEEE transactions on wireless communications, 2020, 19(1): 77–90. DOI: 10.1109/TWC.2019.2941478<http://dx.doi.org/10.1109/TWC.2019.2941478>
- [8] BJÖRNSON E, SANGUINETTI L. Scalable cell-free massive MIMO systems [J]. IEEE transactions on communications, 2020, 68(7): 4247–4261. DOI: 10.1109/TCOMM.2020.2987311<http://dx.doi.org/10.1109/TCOMM.2020.2987311>
- [9] KE M L, GAO Z, WU Y P, et al. Massive access in cell-free massive MIMO-based Internet of Things: cloud computing and edge computing paradigms [J]. IEEE journal on selected areas in communications, 2021, 39(3): 756–772. DOI: 10.1109/JSAC.2020.3018807<http://dx.doi.org/10.1109/JSAC.2020.3018807>
- [10] KE M L, GAO Z, WU Y P, et al. Compressive sensing-based adaptive active user detection and channel estimation: massive access meets massive MIMO [J]. IEEE transactions on signal processing, 2020, 68: 764–779. DOI: 10.1109/TSP.2020.2967175<http://dx.doi.org/10.1109/TSP.2020.2967175>
- [11] GAO Z, DAI L L, WANG Z C, et al. Spatially common sparsity based adaptive channel estimation and feedback for FDD massive MIMO [J]. IEEE transactions on signal processing, 2015, 63(23): 6169–6183. DOI: 10.1109/TSP.2015.2463260<http://dx.doi.org/10.1109/TSP.2015.2463260>
- [12] NAM J, ADHIKARY A, AHN J Y, et al. Joint spatial division and multiplexing: opportunistic beamforming, user grouping and simplified downlink scheduling [J]. IEEE journal of selected topics in signal processing, 2014, 8(5): 876–890. DOI: 10.1109/JSTSP.2014.2313808<http://dx.doi.org/10.1109/JSTSP.2014.2313808>
- [13] WANG D M, YOU X H, HUANG Y M, et al. Full-spectrum cell-free RAN for 6G systems: system design and experimental results [J]. Science China information sciences, 2023, 66(3): 130305. DOI: 10.1007/s11432-022-3664-x<http://dx.doi.org/10.1007/s11432-022-3664-x>
- [14] BJÖRNSON E, JALDEN N, BENGTSSON M, et al. Optimality properties, distributed strategies, and measurement-based evaluation of coordinated multicell OFDMA transmission [J]. IEEE transactions on signal processing, 2011, 59(12): 6086–6101. DOI: 10.1109/TSP.2011.2165706<http://dx.doi.org/10.1109/TSP.2011.2165706>

作者简介



胡彦丰, 东南大学在读博士研究生; 主要研究领域为大规模天线系统的随机接入方案; 先后参与基金项目3项; 已发表论文3篇, 申请专利1项。



王东明, 东南大学教授; 主要研究领域为未来移动通信系统和无线传输技术研究; 先后主持和参与基金项目10余项, 获得科研成果奖2项; 已发表论文100余篇。



梁楚龙, 中兴通讯股份有限公司高级系统工程师; 主要研究领域为信道编码理论及其在通信系统中的应用; 已发表论文20余篇。



尤肖虎, 中国科学院院士、东南大学移动通信国家重点实验室主任; 目前主要研究方向为无线与移动通信系统、现代数字信号处理等; 作为课题负责人, 曾承担10余项国家“863”、科技攻关、国家自然科学基金等项目, 在IEEE Transactions系列期刊上发表论文60余篇, 出版专著2本。

智能增强的免授权多址接入技术



AI-Enhanced Grant-Free Multiple Access Technologies

于含笑/YU Hanxiao, 费泽松/FEI Zesong

(北京理工大学通信技术研究所, 中国 北京 100081)
(Research Institute of Communication Technology, Beijing Institute of Technology, Beijing 100081, China)

DOI: 10.12142/ZTETJ.202401007

网络出版地址: <http://kns.cnki.net/kcms/detail/34.1228.TN.20240227.1008.002.html>

网络出版日期: 2024-02-27

收稿日期: 2024-01-16

摘要: 免授权多址接入技术是未来海量物联 (IoT) 场景的关键使能技术。在免授权接入系统中, 由于用户和基站间缺少授权信息, 易产生用户难以识别以及用户间强干扰等问题, 一方面可将免授权多址技术与非正交多址接入技术结合来缓解随机接入造成的用户干扰问题, 另一方面也可以进一步利用人工智能技术强大的数据特征提取能力, 对收发机算法进行优化, 以有效降低用户碰撞概率, 实现海量物联场景下的大规模通信。分析了当前未来移动通信网络中多址接入技术的需求、发展历程及研究现状。指出未来需要针对全场景进行接入技术拓展和智能技术弹性演进研究, 以助力智能内生多址技术发展。

关键词: 免授权接入; 海量物联; 非正交多址接入; 人工智能

Abstract: The grant-free multiple access technology is regarded as a key enabling technology for future massive Internet of Things (IoT) scenario. In the grant-free access system, the lack of scheduling information between the user and the base station brings the problems such as difficulty to identify the active user set and strong interference between users. Hence, the combination of the grant-free multiple access technology and the non-orthogonal multiple access technology can enhance the user interference elimination capability, and furthermore, the powerful data feature extraction capability of artificial intelligence technology can be utilized to optimize the transceiver algorithms of the authorization-free multiple access system, which can effectively reduce the collision probability of the user and enhance the capability of user identification and collision data recovery. The demand, the development, and the state of art of the multiple access technology in the future wireless communication network are summarized. It is pointed out that in the future, it is necessary to expand access technology for all scenarios and conduct research on the elastic evolution of intelligent technology, in order to support the development of intelligent endogenous multiple access technology.

Keywords: grant-free access; massive IoT; non-orthogonal multiple access; artificial intelligence

引用格式: 于含笑, 费泽松. 智能增强的免授权多址接入技术 [J]. 中兴通讯技术, 2024, 30(1): 33-40. DOI: 10.12142/ZTETJ.202401007

Citation: YU H X, FEI Z S. AI-enhanced grant-free multiple access technologies [J]. ZTE technology journal, 2024, 30(1): 33-40. DOI: 10.12142/ZTETJ.202401007

移动通信技术经历了1G时代到5G时代的演进, 从传统人-人、人-机通信, 逐渐拓展到人-机-物三元间通信。在6G构建及部署初期, 工业自动化、智能电网等垂直物联网应用需要无处不在的无线通信基础设施, 为未来的智能和互联的社会提供全面集成的支持平台。但是, 大多数机器类通信设备具有低计算、小存储以及低功率预算特征, 为了给海量物联网设备提供有效服务, 亟需对未来通信网络系统多址接入技术进行重构, 以满足未来大规模设备接入需求^[1]。

多址接入技术是一种通过对时、频、空、码等无线资源进行划分, 构造用户特有传输通道以实现多用户通信的关键使能技术, 也是移动通信系统持续演进的重要推动力。每一

代通信系统都有自己独特的多址接入技术, 第1代移动通信技术主要采用频分多址接入 (FDMA) 技术, 通过用户的数据在不同的频带上传输而避免用户间信号干扰。第2代移动通信系统采用时分多址接入 (TDMA) 技术, 利用不同时间隙来区分用户。第3代移动通信系统则采用码分多址接入 (CDMA) 技术, 利用不同的正交码字对用户进行区分, 从而在接收端对不同用户进行识别。第4代移动通信系统采用正交频分复用 (OFDM) 技术, 在频分复用的基础上进一步压缩频带, 提高频谱利用率。到了第5代移动通信系统, 为满足物联网场景下大规模连接、低时延、高吞吐量的需求, 学术界开始关注于非正交多址与免授权接入技术。

1 免授权多址接入技术发展及研究现状

基于授权的接入协议要求每一个激活用户都需要获得基站的许可或者授权来接入网络。与之相对,免授权接入技术在用户接入过程中简化了中心授权环节,其核心为:令用户以“到达即转发”的模式随机传输数据,从而达到不发送或发送少量授权信息的目的。由于简化了繁琐的握手协议,显而易见地,免授权接入可以大大节省连接及发送信号所需要的授权开销。同时,由于用户无须等待基站发送许可,因此可以缩短用户的等待时间,降低用户的待机时间,节省用户能耗开销。相较于基于授权的接入技术,免授权接入技术可以有效应对未来大规模物联网通信场景的以下需求:

1) 海量接入终端需求:预计到2025年,物联网终端总数将超过400亿,并于2030年突破1 000亿。面向B5G/6G的海量连接场景,终端数密度将达到千万以上/平方公里。为了应对如此高密度的接入设备,满足终端并发接入需求并简化接入流程成为使能物联通信的关键。

2) 短数据包传输需求:与人-人通信场景下的数据传输特征相比,无线传感终端将呈现小数据包传输以及低数据传输速率特征。物联终端的传输数据包通常在几十上百字节,尤其是抄表、传感、监测信息通常尺寸较小,约为10~20 B。在上述小数据包传输场景中,授权流程中的控制信令所占比重尤为显著,因此对于低开销的接入技术需求更加迫切。

3) 低能耗终端通信需求:为了满足“万物互联”的最终愿景,物联终端势必呈现小型化、广部署的特征,若采用有线供能或者长期频繁为其更换电池,会对系统部署造成极大的负担。因此,需提供低能耗、低成本的通信服务,以满足长生命周期的需求。

尽管免授权技术具有支撑海量设备低时延、低信令开销接入的潜在优势,但是也引入了新的问题:

1) 由于免授权造成基站端缺少终端的先验信息,因此相对于授权场景,需要对激活用户进行检测,此时给用户信道估计也带来了新挑战。

2) 终端数据具有偶发特性,随机的数据传输模式易造成用户数据碰撞,且该碰撞难以避免,若无法应对多用户碰撞问题将导致接入效率下降。

为应对上述问题,将免授权和非正交多址接入(NOMA)技术融合,利用NOMA中多用户叠加传输的特征结合免授权的优势,可以实现对叠加传输用户数据的分离,提升针对用户叠加传输情况的容忍度,为消除用户间干扰提供有效技术途径。多类NOMA技术均针对免授权传输方式进

行了增强设计。例如,多用户共享接入技术(MUSA)具有码本构造简单以及利于部署的特点,可直接推广至免授权系统^[2]。华为设计了免授权稀疏码分多址接入(SCMA)技术,将时间-频率资源划分为竞争传输单元(CTU),用户基于竞争选择竞争传输单元进行数据传输,若碰撞则接收反馈重发^[3]。

上述研究工作提供了初步的免授权NOMA技术的实现思路。然而,由于免授权接入中基站先验信息缺失,系统中接收机面临极大的用户识别负担与数据复杂处理问题,对于随机接入用户的用户集合检测能力和发送数据恢复能力仍显不足。因此,对于免授权下发送端的用户特征码本以及接收端的用户识别、信道估计、用户数据检测算法进行针对性设计与优化显得尤为重要。

随着人工智能技术的不断发展,在处理复杂且动态变化的问题上,机器学习(ML)算法在5G领域的相关研究中得到了广泛的应用。基于能够利用多个工程领域的数据和计算资源的优势,当通信领域知识获取的传统方法中存在建模或算法构建困难时,机器学习可以加快设计周期,降低实施的复杂性和成本,并提高已知算法的性能。机器学习技术的引入可以集成传统的基于模型的工程技术和现有的领域知识,实现两种方法的互补以及协同。截至目前,将机器学习方法应用于免授权接入机制的优化也初显成效,部分学者应用深度学习优化了免授权接入系统的发送码本和接收机。

在文献[4]中,作者提出了一种基于深度学习的非正交随机接入方法,利用相同前导码的多个节点在相同的时频资源上同时发送数据,并基于无监督深度学习进行功率控制。文献[5]提出了一种基于深度学习的端到端随机接入框架,使用深度神经网络实现对冲突前导码的检测和解析,以提高随机接入成功的概率,同时减少整个随机接入过程的延迟。文献[6]面向mMTC上行链路框架中的免授权SCMA系统,提出了单基站下基于组的深度神经网络激活用户检测方案。在文献[7]中,作者在免授权NOMA系统中提出了基于深度学习的用户激活检测和信道估计方案,该方案采用长短期记忆(LSTM)网络计算出基站接收到的NOMA信号与激活设备和信道索引间的映射关系,从而增强激活用户集合与信道检测性能。在文献[8]中,作者解决了免授权传输中的信道估计和用户激活问题,在深度网络结构中引入稀疏贝叶斯学习算法,增强检测结果的准确性。文献[9]研究了mMTC场景下的基于Q学习的随机接入方法,通过引入设备聚类 and NOMA,在基站上进行自适应帧大小调整,以获得最佳的接入策略。文献[10]提出了基于深度强化学习的多用户分布式频谱访问算法,该算法在一个中心单元中完成深度Q网络(DQN)训

练,然后每个用户应用训练好的DQN,通过观察自身局部状态选择合适的频谱访问动作。基于上述研究工作基础,本论文将对基于深度学习的免授权NOMA系统开展进一步分析与研究,提出智能增强的免授权接入收发机设计方案,提升激活用户集合检测及数据检测准确度。

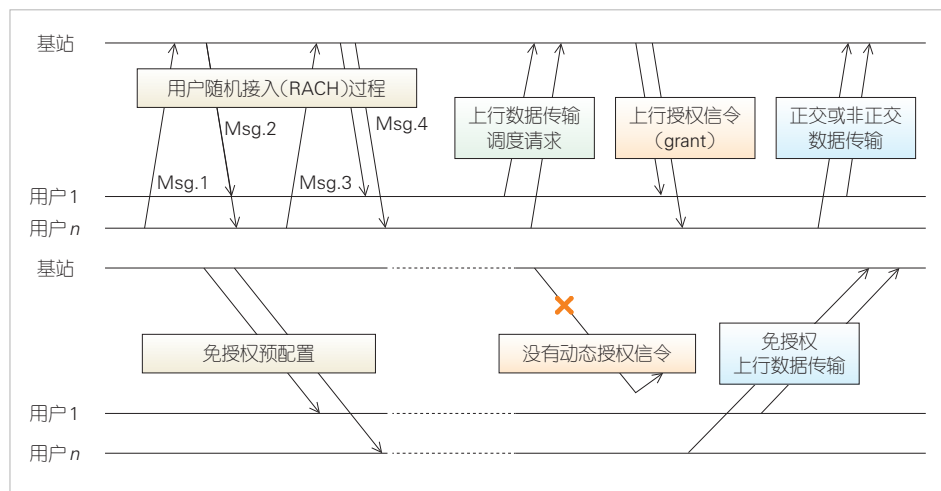
2 免授权多址接入技术系统模型及应用

基于授权的接入协议需要4步握手过程,但是当其运用于大规模终端接入授权通信时,信令开销会变得难以承受,从而对系统造成极大的负担。免授权接入技术的引入可以对该流程进行简化。图1展示了基于授权接入与免授权接入的流程对比。

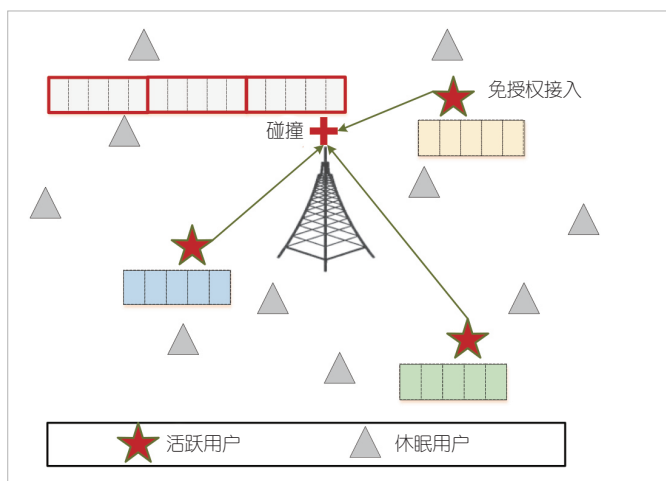
图2为一个典型的上行免授权接入系统,基站位于小区中心,周围用户离散随机分布在小区内,用户数为 N 。用户根据业务需求确认其活跃状态,当数据到达缓冲器时,用户激活并向基站发送传输块,在没有上传数据需求时,用户处于休眠态。在基站与用户以及用户之间保持同步传输的基础上,用户的激活模式可以建模为在给定时间间隔 Δt 内,以 p_n 的概率激活触发数据发送行为。

在给定激活概率后,仍需要定义用户在某一传输时刻的激活行为:以 a_n 表示用户 n 在该时刻的激活状态, $a_n = 1$ 表示用户 n 在该时刻处于发送数据状态;反之, $a_n = 0$ 表示用户 n 在该时刻处于休眠模式。对于激活态的用户,首先生成调制符号 $q_n \in \mathcal{X}_n$,其中 $\mathcal{X}_n$ 为用户 n 的星座图;对于休眠态的用户,可假设其传输信号为零。将激活用户的表示向量为:

$$\gamma = [a_1, \dots, a_n, \dots, a_N], \quad (1)$$



▲图1 授权接入与免授权接入流程对比图



▲图2 上行系统用户随机激活小区示意图

用户 n 的传输符号可以统一表示为:

$$x_n = a_n q_n = \begin{cases} q_n, & a_n = 1 \\ 0, & a_n = 0. \end{cases} \quad (2)$$

假设用户 n 具有特有的扩展序列 $s_n = [s_1, s_2, \dots, s_H]^T \in \mathbb{C}^{H \times 1}$, 序列长度为 H 。用户 n 对其待传输符号 x_n 进行扩展,则扩展后的序列为 $z_n = s_n x_n$ 。方便起见,将 N 个用户的发送符号和发送信号分别用向量 $\mathbf{x}$ 和矩阵 $\mathbf{Z}$ 表示如下:

$$\mathbf{x} = [x_1, x_2, \dots, x_n, \dots, x_N]^T, \quad (3)$$

$$\mathbf{Z} = [z_1^T, z_2^T, \dots, z_n^T, \dots, z_N^T]^T. \quad (4)$$

3 智能辅助的免授权多址接入技术

采用人工智能对免授权多址技术进行优化的方案可大体分为3类:

一类是对发送机和接收机联合设计,也就是采用自编码器优化多用户码本和多用户接收机;一类是对接收机进行设计,包括采用深度学习或卷积神经网络搭建接收机,完成激活用户检测或用户数据检测;第3类为采用强化学习技术优化用户随机发送策略。

3.1 智能多用户码本与接收机设计

在免授权非正交接入系统中,可以考虑由基站预先为用户配置没有冲突的扩展序列。 N 个用户对

的扩展序列集合可表示为：

$$\mathcal{S} = \{s_n\}_{1 \leq n \leq N} \quad (5)$$

在接收到用户的叠加干扰信号后，基站需要根据扩展序列的先验信息，对多用户的信号进行译码。此时，上行传输中用户 n 的信号处理过程可视为从调制符号 x_n 到待传输信号 z_n 的映射过程，其映射方程表示为：

$$z_n = f_n(x_n; \mathbf{s}_n) \quad (6)$$

由于用户间的传输信号互相干扰，因此多用户的扩展序列应联合优化设计。基站的接收信号可以用以下公式表示：

$$\mathbf{y} = \sum_{n=1}^N \text{diag}(\mathbf{h}_n) z_n + \mathbf{n} \quad (7)$$

其中， $\mathbf{y} = [y_1, \dots, y_H]^T$ 。 $\mathbf{n} \sim \mathcal{CN}(0, \sigma^2 \mathbf{I})$ 是方差为 σ^2 的加性高斯白噪声。 $\mathbf{h}_n = [h_{n,1}, \dots, h_{n,H}]^T$ 为用户 n 的信道系数。 $\text{diag}(\mathbf{h}_n)$ 表示为将向量 $\mathbf{h}_n$ 元素扩展为对数矩阵的对角元素的操作。所有用户的信道信息矩阵表示为 $\mathbf{H} = [\mathbf{h}_1, \dots, \mathbf{h}_N]^T$ 。在接收机端，接收到激活用户的叠加信号 $\mathbf{y}$ 后，接收机将针对多用户信号进行联合译码。接收机的信号处理过程也可以用映射函数表示，其方程为：

$$\hat{\mathbf{x}} = g(\mathbf{y}; \mathcal{S}, \mathbf{H}), \quad (8)$$

其中， $\hat{\mathbf{x}} = [\hat{x}_1, \hat{x}_2, \dots, \hat{x}_N]^T$ 为用户符号的估计结果。对系统中用户 n 的通信性能评估标准可设置为估计结果与发送符号的均方误差，也就是：

$$\mathcal{L}_{\text{MSE}, n}(\hat{x}_n, x_n) = \mathbb{E}[\hat{x}_n - x_n]^2 \quad (9)$$

非正交传输系统的优化目标可设置为最小化所有 N 个潜在用户的均方误差和，表示为

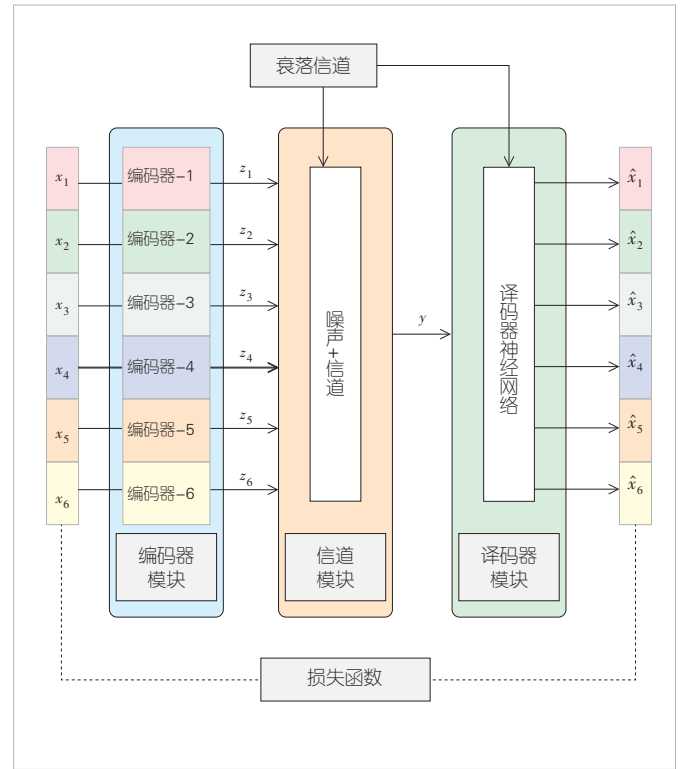
$$\mathcal{L}_{\text{MSE}}(\hat{\mathbf{x}}, \mathbf{x}) = \sum_{n=1}^N \mathcal{L}_{\text{MSE}, n}(\hat{x}_n, x_n) \quad (10)$$

以公式 (10) 作为代价函数优化神经网络，可以得到免授权 NOMA 的扩展序列优化问题：

$$P1: \{S^*, g^*\} = \min_{S = \{s_n\}_{1 \leq n \leq N}, g} \mathcal{L}_{\text{MSE}} \quad (11)$$

为了优化上述系统的码本集合，即 $\mathcal{S} = \{s_n\}_{1 \leq n \leq N}$ ，可以利用自编码器的结构，设计基于自编码器的免授权 NOMA 优化网络，其结构如图 3 所示，具体由 3 部分构成：

1) 编码器：编码器部分由 N 个独立的小模块组成，每个模块模拟了单个用户的信号产生及发送过程。每个模块为一个小规模神经网络。第 n ($n \in \{1, 2, \dots, N\}$) 个模块输入为用户 n 的调制符号 x_n ，并实现将 x_n 扩频为待发送信号 z_n 的功



▲图3 基于自编码器的多址接入优化

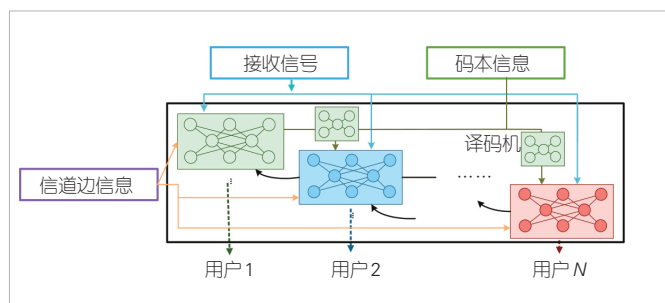
能。神经网络参数充当了扩展序列，实现对符号扩展。图 3 中， N 个模块独立地实现对 N 个用户的符号扩展处理。

2) 信道层：在信道层中，编码器产生的信号受到信道衰落 $\mathbf{h}$ 和加性高斯白噪声 $\mathbf{n}$ 的影响。

3) 译码器：译码器部分模拟基站中的译码过程，探索叠加传输信号的内部特征，从而恢复用户信号。译码器结构将直接影响扩展序列的最终优化性能。

编码器和译码器在神经网络训练过程中联合训练，可同时优化免授权 NOMA 中的多用户扩展序列及多用户接收算法。在前向传播阶段，数据流依次通过编码器、信道层和译码器后，输出受到信道和噪声污染的发送信号的估计信号。接着，设置代价函数衡量估计信号的准确性，在反向传播过程中，神经网络的参数以最小化代价函数的目标通过梯度下降法更新。

受串行干扰消除接收机的启发，我们提出了仿串行干扰消除的多用户神经网络接收机结构。如图 4 所示，在接收机结构设计过程引入串行干扰消除思想，利用用户信道边信息，相对于全连接网络接收机，在可控复杂度条件下能有效提升接收机译码性能。所提接收机由 N 个相同结构的模块构成，第 n 个模块输入为接收信号和用户信道信息，输出为用户 n 的译码结果。每个模块与其他模块依次连接，形成串联



▲图4 仿串行干扰消除的多用户深度学习接收机示意图

关系，引入上一模块的估计结果，串行地译码多用户信号，模块间的信号处理过程将互相影响，得到更准确的多用户信号检测结果。本文提出的多用户神经网络接收机，相对于全连接网络接收机，在可控复杂度条件下能有效提升接收机译码性能。

此外，本文在优化中可以进一步考虑用户公平性，例如对于激活概率具有明显差异的用户进行性能加权评估，加速神经网络训练过程，使其趋向于更快地优化高激活概率用户的扩展序列。此时，可以在设计代价函数时引入公平项代价，对多用户的均方误差均衡度进行评估，定义如下：

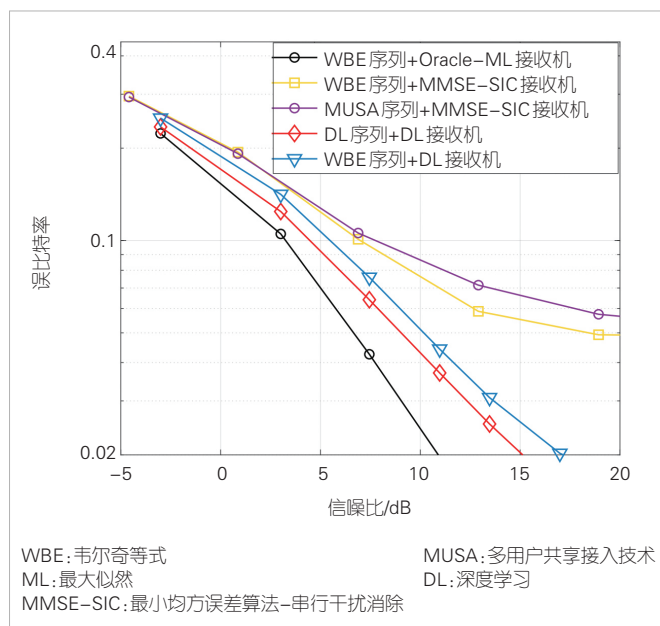
$$\mathcal{L}_{\text{fair}}(\hat{\mathbf{x}}, \mathbf{x}) = \sum_{n=1}^N \lambda_n |\mathcal{L}_{\text{MSE}, n} - \mathcal{L}_{\text{MSE}}|, \quad (12)$$

其中， λ_n 为用户 n 的权重。基于此，总体系统的代价函数设置为公平项代价与均分误差代价的加权和，即

$$\mathcal{L}^{\text{ALL}}(\hat{\mathbf{x}}, \mathbf{x}) = \beta_{\text{MSE}} \mathcal{L}_{\text{MSE}}(\hat{\mathbf{x}}, \mathbf{x}) + \beta_{\text{fair}} \mathcal{L}_{\text{fair}}(\hat{\mathbf{x}}, \mathbf{x}), \quad (13)$$

其中， β_{MSE} 和 β_{fair} 为对应的权重系数。用公式(13)代替公式(10)作为神经网络代价函数，可以均衡用户公平性与系统总体译码性能。

为了满足实际系统需求，本文所提方案对于代价函数进行优化，并在编码器和译码器的设计中嵌入原有通信模型，可以加速神经网络训练，提升译码结果准确性。图5仿真对比了本文所提的智能优化的非正交多址收发机与传统非正交多址收发机的性能。仿真中对比了所提深度学习序列与传统MUSA序列和WBE (Welch - Bound Equality) 序列的误码率性能，并对比了接收机与传统MMSE - SIC接收机和最大似然 (ML) 接收机的性能。仿真图中，性能最优的黑色线为假设免授权系统中激活用户集合在接收机侧已知情况下的接收机性能。从图5中可以观察得到，在应用传统WBE序列场景时，应用所提深度学习接收机性能优于传统接收机，此外，应用联合的智能收发机可获得最明显的误比特率增益。



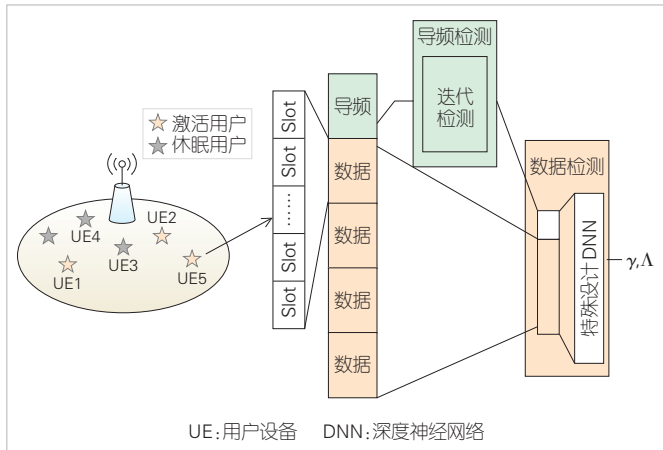
▲图5 智能免授权码本接收机与传统收发算法性能对比

3.2 激活用户集合的智能检测接收

上一节提出的收发机仅解决了用户数据收发问题，但免授权接入系统中基站还需要检测激活用户集合以及激活用户信道状态。本节将设计智能接收机解决非正交免授权接入系统中的联合激活用户集合检测和用户信道估计问题。不同的数据发送模式将影响求解激活用户检测与信道估计问题的难度。纯数据模式传输模型中，用户仅仅发送数据信号，基站根据叠加的用户数据信号进行激活用户检测、信道估计以及数据信号译码。此种方案将数据传输的额外信令开销降到了极限，然而却对接收端提出了更高的要求。在更普遍的免授权传输方案中，为均衡用户开销和传输复杂度，用户发送非正交导频作为信道估计与用户识别的标志，但非正交导频间的用户干扰也为接收机的检测带来新挑战。

面对较短的导频序列带来的检测准确度不足的问题，考虑到数据信号中也包含用户激活信息，因此挖掘数据信号内在特征辅助激活用户集合检测，可获得显著的检测准确度增益。图6展示了一种免授权接入系统下的接收端智能检测接收机设计方法，在发送端随机激活用户首先发送特有导频信号作为自身标志，接着利用特有码本作为数据扩展序列并无间断发送多个数据信号。接收端通过搭建智能接收机对导频信号和数据信号进行联合估计，输出激活用户集合检测结果。

本节仍针对免授权非正交多址接入系统，用户为单天线，基站为多天线或单天线。用户随机激活，接着向基站发



▲图6 数据辅助导频的联合激活用户检测与信道估计

送导频信号和数据信号。其中，导频信号为用户特有的，且和用户自身ID对应；数据信号为用户对发送符号进行扩展生成的。用户每次数据发送中，发送的传输块包括导频信号 y_p 和 M 个数据信号 $y_{[m]}, m = 1, \dots, M$ 。接收端接收到导频信号及数据信号。接着，检测活跃用户集合和激活用户信道状态。定义 $\mathcal{S}$ 用户特有导频序列集合， $\mathbf{x}_{[n]}$ 为用户 n 的发送数据信号， $\hat{\gamma}$ 和 $\hat{\Lambda}$ 分别表示检测的激活用户集合及信道估计结果。将接收到的数据信号和导频信号联合表示为 $\mathbf{Y}$ ，数据信号辅助的联合激活用户检测及信道估计问题可以表示为：

$$\{\hat{\gamma}, \hat{\Lambda}\} = \underset{\gamma, \Lambda}{\operatorname{argmax}} \sum_{[\mathbf{x}_{[n]}]_{n=1}^N} P(\mathbf{Y} | \gamma, \Lambda, \mathcal{S}, [\mathbf{x}_{[n]}]_{n=1}^N). \quad (14)$$

在上行免授权非正交多址接入系统中，为了解决公式 (14) 提出的联合用户活跃检测和信道估计问题，本文设计用户激活检测与信道估计网络 (UAD-CE-NN)。UAD-CE-NN 网络部署于基站，在基站接收到一个传输块后进行用户活跃检测和信道估计，由导频检测神经网络与数据检测神经网络两个模块构成。这两个模块分别对用户发送的导频信号与数据信号进行挖掘，充分提取不同信号中的信息。UAD-CE-NN 对接收信号的处理过程可以由以下映射函数表示：

$$\hat{\Lambda} = f(\mathbf{Y}; \mathcal{S}). \quad (15)$$

由于激活用户检测与其信道估计结果具有强相关性，当用户处于非激活状态时，信道估计结果为0。因此，在解决联合激活用户与信道检测问题时，可以仅以信道估计结果为输出，接着通过 sgn 转换函数实现从信道估计结果到用户激活检测结果的转换。该操作表示为：

$$a_n = [\eta(\hat{\Lambda}; \lambda)]_n = \operatorname{sgn}(\|\hat{\Lambda}_n\|_2^2 - \lambda), \quad (16)$$

其中， λ 为可调的辅助参数用于平衡虚警和误检概率。相对

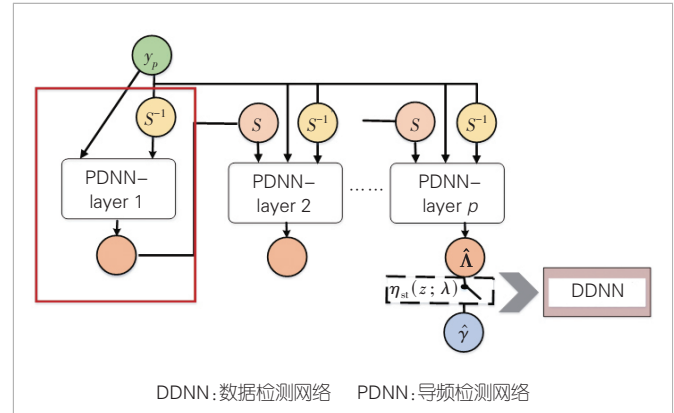
于传统压缩感知算法，智能检测接收机可以显著提升信道估计和激活用户集合检测准确度。此外，相对于仅仅检测导频的深度神经网络算法，智能检测接收机也可获得较好的检测性能。

如图7所示的导频检测网络 (PDNN)，输入信号为基站接收到的导频信号，输出初步活跃用户集与信道估计结果。PDNN的设计引入了仿压缩感知算法的迭代结构，即通过设计 p 个串行的相似神经网络单元，依次对上一单元输出结果进行增强检测。每个PDNN内的神经网络单元为一个PDNN-layer，其输入为接收信号或上一模块估计结果的残差，输出为信道估计结果。在最终单元输出信道估计结果后，转换函数将信道估计结果转换为活跃用户检测结果。最终的输出结果表示：

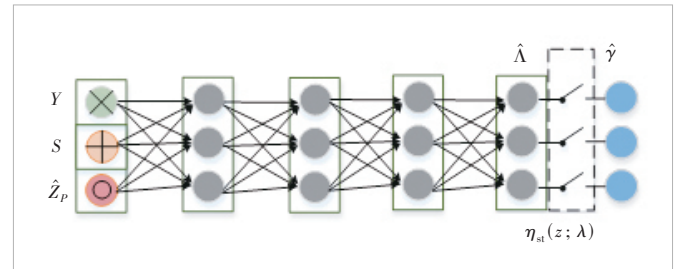
$$\hat{\Lambda}_p = f_p(y_p; \mathcal{S}), \quad (17)$$

$$\hat{\gamma}_p = \eta_{\text{st}}(\hat{\Lambda}_p; \lambda_p). \quad (18)$$

如图8所示的数据检测网络 (DDNN) 通过深度神经网络对数据中包含信息进行挖掘，与导频信号中信息结合，共同完成活跃用户和信道状态检测。PDNN对导频信号进行处理，实现初步估计。DDNN级联于PDNN之后，以接收到的数据信号、导频信号和PDNN检测结果为输入，通过进一步探索数据信号信息，对PDNN输出结果进行纠正，提升输出



▲图7 导频检测网络示意图



▲图8 数据检测网络示意图

结果可靠性。与PDNN类似，DDNN数据处理过程可表示为以下映射方程：

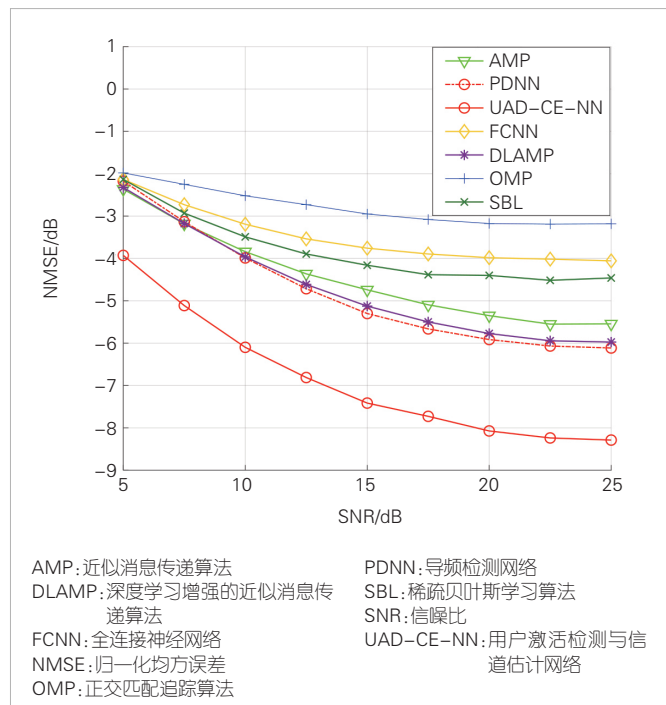
$$\hat{\Lambda}_D = f_D(\mathbf{y}_p, [\mathbf{y}_{[m]}]_{m=1}^M; \mathcal{S}, \hat{\Lambda}_p). \quad (19)$$

我们在瑞利信道和莱斯信道场景下对所提方案进行了性能仿真验证。图9仿真中展示了在只使用前导序列进行联合用户活动检测和信道估计的情况下，不同信噪比的信道估计的归一化均方误差（NMSE）。除了所提方法，我们还展示了传统压缩传感方法的性能及全连接神经网络方案的性能。我们构建了与所提网络计算复杂度类似的全连接神经网络方案，输入与所提网络相同。仿真结果表明，尽管只利用了导频信号来估计信道状态，相比于基于压缩感知的方法，PDNN仍取得了性能上的提升。这说明了所提方案在利用数据信号中包含的信息方面的效率。此外，相对于全连接神经网络，本文所提模型驱动的神经网络检测精度明显提升。

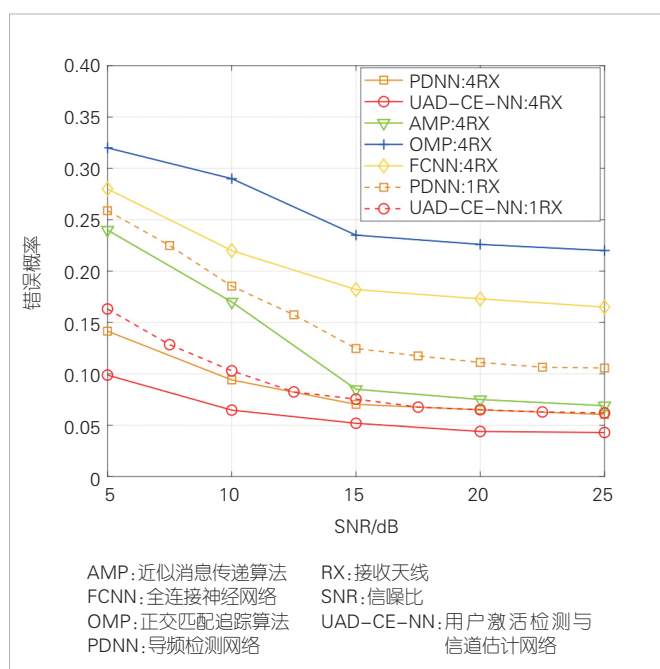
在图10中，我们显示了基站配备单天线和4天线场景下，仅利用前导序列时检测活跃用户状态错误概率与SNR的关系。与图9中的结果类似，我们所提出的PDNN和UAD-CE-NN在用户活动检测方面优于其他方案。仿真结果验证了所提出的UAD-CE-NN相比传统方案的性能增益。

4 结束语

免授权多址接入技术是使能大规模物联终端低时延、低



▲图9 信道估计结果与基准方案性能对比示意图



▲图10 用户激活状态性能对比示意图

开销接入的有效技术手段。面向免授权接入过程导致的用户数据叠加干扰问题，通过与非正交多址技术融合，并利用人工智能对免授权多址收发机进行增强设计，基于数据挖掘能力与通信模型的联合驱动，可有效提升数据传输可靠性以及频谱资源利用率。面对未来6G在无线通信场景中所提出的新愿景以及更高指标需求，我们仍需针对全场景进行接入技术的拓展并开展智能技术的弹性演进研究，对无源场景、感知业务场景、非地面等新场景完成针对性的增强设计及协议研发，并在智能增强方面实现鲁棒性增强、自演进能力进一步拓展，以助力智能内生多址技术的发展。

参考文献

- [1] IMT-2030(6G)推进组. 6G典型场景和关键能力 [R]. 2022
- [2] YUAN Z F, YU G H, LI W M, et al. Multi-user shared access for Internet of Things [C]//Proceedings of IEEE 83rd Vehicular Technology Conference (VTC Spring). IEEE, 2016: 1-5. DOI: 10.1109/VTCSpring.2016.7504361
- [3] AU K, ZHANG L Q, NIKOPOUR H, et al. Uplink contention based SCMA for 5G radio access [C]//Proceedings of IEEE Globecom Workshops (GC Wkshps). IEEE, 2014: 900-905. DOI: 10.1109/GLOCOMW.2014.7063547
- [4] JANG H S, LEE H, QUEK T Q S, et al. Deep learning-based cellular random access framework [J]. IEEE transactions on wireless communications, 2021, 20(11): 7503-7518. DOI: 10.1109/TWC.2021.3085303
- [5] NASIR Y S, GUO D N. Multi-agent deep reinforcement learning for dynamic power allocation in wireless networks [J]. IEEE

- journal on selected areas in communications, 2019, 37(10): 2239–2250. DOI: 10.1109/JSAC.2019.2933973
- [6] HE H T, WEN C K, JIN S, et al. Deep learning-based channel estimation for beamspace mmWave massive MIMO systems [J]. IEEE wireless communications letters, 2018, 7(5): 852–855. DOI: 10.1109/LWC.2018.2832128
- [7] AHN Y, KIM W, SHIM B. Active user detection and channel estimation for massive machine-type communication: deep learning approach [J]. IEEE Internet of Things journal, 2022, 9(14): 11904–11917. DOI: 10.1109/JIOT.2021.3132329
- [8] DING T, YUAN X J, LIEW S C. Sparsity learning-based multiuser detection in grant-free massive-device multiple access [J]. IEEE transactions on wireless communications, 2019, 18(7): 3569–3582. DOI: 10.1109/TWC.2019.2915955
- [9] LIU J J, SHI Z J, ZHANG S W, et al. Distributed Q-learning aided uplink grant-free NOMA for massive machine-type communications [J]. IEEE journal on selected areas in communications, 2021, 39(7): 2029–2041. DOI: 10.1109/JSAC.2021.3078496
- [10] NAPARSTEK O, COHEN K. Deep multi-user reinforcement learning for distributed dynamic spectrum access [J]. IEEE transactions on wireless communications, 2019, 18(1): 310–323. DOI: 10.1109/TWC.2018.2879433

作者简介



于含笑，北京理工大学通信技术研究所博士后；主要研究领域为5G/6G通信网络、低轨卫星通信网络；主持并参与国家自然科学基金青年、国家重点研发等6项；发表论文12篇。



费泽松，北京理工大学通信技术研究所教授；主要研究领域为5G/6G通信网络、通感一体化网络及空天地一体化网络；先后主持国家自然科学基金重点与面上项目、国家科技重大专项以及国家重点研发计划等10余项国家级项目，获得3项省部级自然科学奖；发表论文100余篇。

共享码本随机接入 有限块长信息理论极限分析



Fundamental Limits of Massive Random Access with A Common Codebook and Finite Blocklength

高俊园/GAO Junyuan, 吴泳澎/WU Yongpeng,
张文军/ZHANG Wenjun

(上海交通大学, 中国 上海 200240)
(Shanghai Jiao Tong University, Shanghai 200240, China)

DOI: 10.12142/ZTETJ.202401008

网络出版地址: <http://kns.cnki.net/kcms/detail/34.1228.tn.20240226.1131.006.html>

网络出版日期: 2024-02-26

收稿日期: 2024-01-08

摘要: 超大规模机器通信 (umMTC) 是 6G 移动通信网络的典型应用场景。针对 umMTC 场景中的共享码本随机接入问题展开信息理论研究, 在给定基站天线数、码长、用户载荷和发送功率的情况下, 推导了平均每用户误差概率的非渐近可达界和逆定理。仿真结果证实了可达界和逆定理的紧致性, 并且证明了当用户数低于某一门限时存在多用户干扰消除效应。基站配置多天线能够显著降低平均每用户误差概率, 这对支持大量用户可靠接入具有重要意义。此外, 与独立码本随机接入相比, 共享码本随机接入在能效性方面更具优势。

关键词: 共享码本; 大规模随机接入; 有限块长; 可达界和逆定理

Abstract: Ultra-massive Machine Type Communication (umMTC) is expected to be a key use case in the 6th Generation wireless networks. The information-theoretical analysis for unsourced random access in umMTC scenarios is conducted. Thenon-asymptotic achievability and converse bounds are established on the per-user probability of error for given number of receive antennas, blocklength, payload size, and transmitting power. Simulation results verify the tightness of our bounds and demonstrate that when the number of users is below a threshold, the multi-user interference can be almost perfectly canceled. The per-user probability of error decreases as the number of receive antennas increases, highlighting the crucial role of MIMO in supporting reliable access for a large number of users. Moreover, it is shown that in contrast to random access with individual codebooks, random access with a common codebook achieves higher energy efficiency.

Keywords: common codebook; massive random access; finite blocklength; achievability and converse bounds

引用格式: 高俊园, 吴泳澎, 张文军. 共享码本随机接入有限块长信息理论极限分析 [J]. 中兴通讯技术, 2024, 30(1): 41-45. DOI: 10.12142/ZTETJ.202401008

Citation: GAO J Y, WU Y P, ZHANG W J. Fundamental limits of massive random access with a common codebook and finite blocklength [J]. ZTE technology journal, 2024, 30(1): 41-45. DOI: 10.12142/ZTETJ.202401008

随着物联网 (IoT) 的迅猛发展, 移动数据流量和物联网设备连接数量呈现出快速增长的趋势。IMT-2030 (6G) 推进组在《6G 典型场景和关键能力》白皮书中, 将超大规模机器通信 (umMTC) 列为 6G 的五大典型场景之一。与传统的多用户接入模型相比, 面向 umMTC 的大规模随机接入场景具有下述 4 个典型特点:

1) 海量的用户与设备接入。网络中有海量的潜在设备

是大规模随机接入场景最显著的特点。6G 关键性能指标为实现每平方千米 1 000 万设备的连接数密度。

2) 设备零星活跃进行数据传输。在 IoT 应用场景中, 数据的传输通常不具有连续性, 同一时刻只有零星的设备被激活进行数据传输。

3) 设备传输的数据量较小。在 umMTC 场景中, 终端通常为警报器、传感器等设备。当这些设备活跃时, 传输的多为传感信息以及控制信息等小数据包。

4) 设备具有低功耗需求。在 umMTC 场景中, 终端设备一般具有小型化和低成本的特点。例如, 在环境监测等应用场景中, 传感器往往部署于山区、深水等复杂环境中, 因

基金项目: 国家重点研发计划 (2018YFB1801102); 国家自然科学基金项目 (62122052、62071289); 中央高校基本科研业务费专项资金项目; 高等学校学科创新引智计划 (BP0719010); 上海市自然科学基金项目 (22DZ2229005)

此,设备通常具有较高的低功耗需求,从而保证超长的供电寿命以避免频繁的电池更换。

上述大规模随机接入场景的特点给现有移动通信系统的设计带来巨大的挑战。为此,全球学者针对大规模随机接入传输理论方法展开了深入研究。根据发送端码本的分配情况,随机接入可分为独立码本随机接入和共享码本随机接入两大类。在独立码本随机接入场景中,每个用户有一个独一无二的码本,接收端可以根据码本信息恢复出用户标识。共享码本随机接入由美国麻省理工大学的Y. POLYANSKIY教授于2017年提出^[1]。当所有用户共享同一个码本时,可能会有码字冲突,但不需要为每个用户设计和分配独立的码本,这样可以避免资源浪费。在这种情况下,接收端只需要检测发送的码字集合,而不需要还原用户的标识。由于搜索的码字空间更小,因此可以实现比独立码本随机接入更高的频谱利用效率。如果需要的话,用户可以通过发送的信息来承载该用户的标识。接收端在恢复码字集合后,就可以根据承载的信息识别用户身份。

考虑到共享码本随机接入在避免码本分配资源损耗以及提升频谱利用效率方面的优势,针对该场景展开信息理论研究对于揭示umMTC场景的理论极限具有重要意义。假设活跃用户数已知,文献[1]针对加性高斯白噪声(AWGN)信道,推导了用户共享码本随机接入时平均每用户误差概率(PUPE)的非渐近可达界。针对信道状态信息未知的单天线瑞利衰落信道模型,文献[2]和[3]基于PUPE准则,推导了为支持每个活跃用户可靠传输 J 比特信息所需每比特能量的非渐近可达界和逆定理。文献[4]进一步考虑了活跃用户数未知的情况,定义了基于平均每用户漏检概率和虚警概率约束的大规模随机接入码,并推导了AWGN信道中共享码本随机接入时的非渐近可达界。在码长 n 趋于无穷的渐近情况下,文献[5]针对AWGN信道中的共享码本随机接入问题,推导了可支持的活跃用户数的渐近标度律。

此外,多输入多输出(MIMO)技术通过配置大量天线来同时服务多个用户,能够显著提升共享码本随机接入的频谱利用效率,进一步增强网络覆盖和系统容量。具体地,假设活跃用户数事先已知但信道状态信息未知,在码长 n 和基站天线数 L 趋于无穷的渐近情况下,文献[6]证明频谱利用效率的数量级能够达到 $O(n \ln n)$;当采用级联编码降低译码复杂度后,频谱利用效率的数量级可达 $O(n/\ln n)$ 。考虑到MIMO技术能够显著提升系统的频谱利用效率,探索MIMO信道中共享码本随机接入的信息理论极限对万物互联有着重要意义。

1 系统模型

考虑一般性的上行共享码本大规模随机接入系统,系统中的总用户数为 K ,其中活跃用户数为 K_a ,活跃用户集合记为 $\mathcal{K}_a$ 。假设所有用户共享同一个码本,该码本包含 $M = 2^J$ 个长度为 n 的码字,将码本矩阵记为 $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_M] \in \mathbb{C}^{n \times M}$ 。每个码字满足最大功率约束,即码字总能量不超过 nP 。令 $S_e = \frac{K_a J}{n}$ 表示频谱利用效率, $E_b = \frac{nP}{J}$ 表示每比特能量。

考虑MIMO准静态瑞利衰落信道模型,基站配置的天线数为 L ,用户配置单天线。假设信道状态信息未知,即接收端不知道信道衰落系数但知道其分布。基站第 l 个天线的接收信号可表示为:

$$\mathbf{y}_l = \sum_{k \in \mathcal{K}_a} h_{k,l} \mathbf{x}_{w_k} + \mathbf{z}_l \in \mathbb{C}^n, \quad (1)$$

其中, $h_{k,l} \sim \mathcal{CN}(0,1)$ 表示用户 k 和基站第 l 个接收天线之间的信道衰落系数,不同的用户和不同的接收天线之间的信道衰落系数满足独立同分布; $\mathbf{z}_l \sim \mathcal{CN}(0, \mathbf{I}_n)$ 表示加性高斯白噪声; w_k 表示活跃用户 k 发送的信息,在集合 $[M]$ 中服从均匀分布,对应的码字记为 $\mathbf{x}_{w_k}$ 。 L 个天线上的接收信号 $\mathbf{Y} = [\mathbf{y}_1, \dots, \mathbf{y}_L]$ 可表示为:

$$\mathbf{Y} = \mathbf{X} \Phi \mathbf{H} + \mathbf{Z} \in \mathbb{C}^{n \times L}, \quad (2)$$

其中,信道系数矩阵 $\mathbf{H} = [\mathbf{h}_1, \dots, \mathbf{h}_L] \in \mathbb{C}^{K \times L}$,对于任意 $l \in [L]$, $\mathbf{h}_l = [h_{1,l}, \dots, h_{K,l}]^T \in \mathbb{C}^K$;噪声矩阵 $\mathbf{Z} = [\mathbf{z}_1, \dots, \mathbf{z}_L] \in \mathbb{C}^{n \times L}$;二元选择矩阵 $\Phi \in \{0,1\}^{M \times K}$ 表征用户发送的信息,具体而言,如果第 k 个用户活跃并发送第 w_k 个码字时, $[\Phi]_{w_k,k} = 1$,否则 $[\Phi]_{w_k,k} = 0$ 。

接收端基于接收信号进行译码,将译出信息的集合记为 $\hat{\mathcal{W}}$,满足 $|\hat{\mathcal{W}}| = K_a$ 。此时,平均每用户误差概率PUPE可表示为:

$$P_e = \frac{1}{K_a} \sum_{k \in \mathcal{K}_a} \mathbb{P} \left[\{w_k \notin \hat{\mathcal{W}}\} \cup \{w_k = w_i \text{ for } i \neq k\} \right]. \quad (3)$$

2 非渐近可达界

在接收端信道状态信息未知的情况下,定理1给出了MIMO准静态瑞利衰落信道中共享码本随机接入时PUPE的可达界。

定理1:假设 K 个潜在用户中共有 K_a 个活跃用户,用户均配置单天线,基站的天线数为 L 。所有用户共享一个码本,包含 $M = 2^J$ 个长度 n 的码字,满足最大功率不超过 nP

约束。在MIMO准静态瑞利衰落信道中,当信道状态信息未知且活跃用户数 K_a 确定已知时,系统的PUPE的上界为:

$$P_e \leq \min_{0 < P' < P} \left\{ p_0 + \sum_{t=1}^{K_a} \frac{t}{K_a} \min\{1, p_t\} \right\}, \quad (4)$$

其中,

$$p_0 = \frac{\binom{K_a}{2}}{M} + K_a \left[1 - \frac{\gamma\left(n, \frac{nP}{P'}\right)}{\Gamma(n)} \right], \quad (5)$$

$$p_t = \min_{0 \leq \omega \leq 1, 0 \leq \nu} \{q_{1,t}(\omega, \nu) + q_{2,t}(\omega, \nu)\}, \quad (6)$$

$$q_{1,t}(\omega, \nu) = \left(\frac{K_a}{t} \right) \binom{M - K_a}{t} \mathbb{E}_{\mathbf{C}} \left[\min_{\substack{u \geq 0, r \geq 0, \\ \lambda_{\min}(\mathbf{B}) > 0}} \exp\{Lr\nu\} \exp\left\{L\left((u - r)\ln|\mathbf{F}| - u\ln|\mathbf{F}_2| + r\omega\ln|\mathbf{F}_1| - \ln|\mathbf{B}|\right)\right\} \right], \quad (7)$$

$$q_{2,t}(\omega, \nu) = \min_{\delta \geq 0} \left(\frac{K_a}{t} \right) \mathbb{E}_{\mathbf{C}} \left[\frac{\gamma(Lm, c_\delta)}{\Gamma(Lm)} + 1 - \frac{\gamma(nL, nL(1 + \delta))}{\Gamma(nL)} \right], \quad (8)$$

$$c_\delta = \frac{L(n(1 + \delta)(1 - \omega) - \omega\ln|\mathbf{F}_1| + \ln|\mathbf{F}| - n\nu)}{\omega \prod_{i=1}^m \lambda_i^{\frac{1}{t}}}, \quad (9)$$

$$\mathbf{B} = (1 - u + r)\mathbf{I}_n + u\mathbf{F}_2^{-1}\mathbf{F} - r\omega\mathbf{F}_1^{-1}\mathbf{F}, \quad (10)$$

$$\mathbf{F} = \mathbf{I}_n + \mathbf{C}\mathbf{F}_{S_{K_a}}\mathbf{C}^H, \quad (11)$$

$$\mathbf{F}_1 = \mathbf{I}_n + \mathbf{C}\mathbf{F}_{S_{K_a} \setminus S_1}\mathbf{C}^H, \quad (12)$$

$$\mathbf{F}_2 = \mathbf{I}_n + \mathbf{C}\mathbf{F}_{S_{K_a} \setminus S_1 \cup S_2}\mathbf{C}^H, \quad (13)$$

其中, S_{K_a} 表示发送信息的集合, 包含 K_a 个元素; S_1 表示集合 S_{K_a} 的任意一个包含 t 个元素的子集; S_2 表示集合 $[M] \setminus S_{K_a}$ 的任意一个包含 t 个元素的子集; 矩阵 $\mathbf{C} \in \mathbb{C}^{n \times M}$ 表示假设无功率约束时的码本矩阵, 其元素服从 i.i.d. $\mathcal{CN}(0, P')$ 分布; 对于任意集合 $S \subset [M]$, 对角矩阵 $\mathbf{F}_S = \text{diag}\{\gamma_s\} \in \{0, 1\}^{M \times M}$ 满足当 $i \in S$ 时, $[\gamma_s]_i = 1$, 否则 $[\gamma_s]_i = 0$; $\lambda_1, \dots, \lambda_n$ 表示矩阵 $\mathbf{F}_1^{-1}\mathbf{C}\mathbf{F}_{S_1}\mathbf{C}^H$ 的特征值, 秩为 $m = \min\{n, t\}$; $\gamma(\dots)$ 表示下不完全 Gamma 函数; $\Gamma(\cdot)$ 表示 Gamma 函数。

证明: 上述定理主要采用高斯随机编码^[7]和最大似然 (ML) 解码方案得到。首先, 考虑一个新的测度, 假设码本中的每一个元素服从独立同分布的 $\mathcal{CN}(0, P')$ 分布, 并假设既不存在码字冲突也没有功率约束。 p_0 表示真实测度与新测度之间全变差距离的上界。接下来, 基于新的测度, 推导出恰好有 t 个码字错误检测的概率的上界, 将该上界记为 p_t 。该上界主要通过 Fano 提出的基于“良好区域”的技巧^[8]、Chernoff 界以及二次型的矩生成函数等数学工具得到。 p_t 主要包含两部分, 第一项 $q_{1,t}(\omega, \nu)$ 表示有恰好 t 个码字漏检且接收信号位于发送信号线性组合周围某一区域的概率的上界, 第二项 $q_{2,t}(\omega, \nu)$ 表示接收信号落在该区域外的概率的上界。

定理 1 给出了 MIMO 准静态瑞利衰落信道中共享码本随机接入时平均每用户误差概率的可达界。在信道状态信息未知的情况下, 信道衰落和噪声的影响会耦合在一起, 在分析中很难将这两种影响分开。幸运的是, 当考虑瑞利衰落信道时, 在给定发送信号的情况下, 接收到的信号 $\mathbf{Y}$ 服从高斯分布, 这使得分析更加容易。当信道服从更加复杂的信道模型时^[9], 系统的性能极限有待进一步探索。此外, 从定理 1 可以看出, 当接收端天线数增加时, 恰好有 t 个码字错误检测的概率的上界 p_t 逐渐减小, 因此, MIMO 技术对提升系统的可靠性具有显著增益。

3 非渐近逆定理

定理 2 给出了在信道状态信息未知情况下, MIMO 准静态瑞利块衰落信道中大规模随机接入时 PUPE 的逆定理。

定理 2: 假设活跃用户数为 K_a , 用户均配置单天线, 基站的天线数为 L 。所有用户共享同一个码本, 包含 $M = 2^J$ 个长度 n 的码字。在信道状态信息未知的 MIMO 准静态瑞利衰落信道中, 当 K_a 确定且已知时, 平均每用户误差概率 P_e 需满足以下条件:

1) 当码本中的元素服从 i.i.d. $\mathcal{CN}(0, P)$ 分布时, 则应满足:

$$nL\log_2(1 + K_a P) - L \left[1 - \frac{\binom{K_a}{2}}{M} \right] \mathbb{E} \left[\log_2 |\mathbf{I}_n + \mathbf{X}_{K_a} \mathbf{X}_{K_a}^H| \right] \leq (1 - P_e) K_a (J - \log_2 K_a) - K_a H_2(P_e), \quad (14)$$

$$P_e \leq \frac{M}{M + K_a}, \quad (15)$$

其中, 矩阵 $\mathbf{X}_{K_a} \in \mathbb{C}^{n \times K_a}$ 中的元素服从 i.i.d. $\mathcal{CN}(0, P)$ 分布。

2) 单用户有限码长理论界表明:

$$J - \log_2 K_a \leq -\log_2 P[\chi^2(2L) \geq (1 + (n+1)P)r], \quad (16)$$

其中, r 是满足以下条件的解:

$$P[\chi^2(2L) \leq r] = P_e. \quad (17)$$

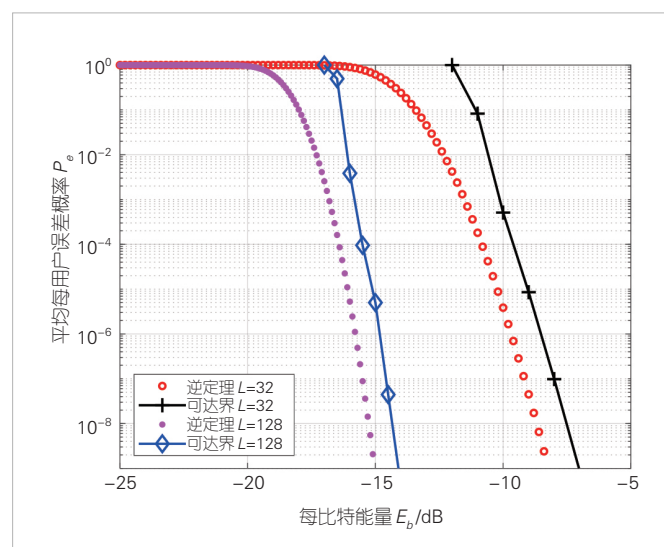
证明: 上述定理中的条件一考虑多用户共享码本随机接入场景, 主要利用 Fano 不等式得到, 其中的互信息依照文献[10]中的方法进行处理。条件二的推导思路如下: 假设有 K_a 个活跃用户, 当 $K_a - 1$ 个活跃用户的信道衰落系数和发送的信息已知时, 检测 K_a 个用户发送信息的问题可以退化成检测单用户发送信息的问题。在共享码本随机接入场景中可能存在信息冲突, 因此, $K_a - 1$ 个活跃用户发送的信息中不同信息的数量 B 满足 $1 \leq B \leq K_a - 1$, 解码器的目标是输出另外 $K_a - B$ 个信息以恢复最后一个活跃用户发送的信息。为了得到逆定理, 将译码列表大小从 $K_a - B$ 扩大到 K_a , 将 PUPE 缩放为 $\mathbb{P}[W_1 \notin \hat{\mathcal{W}}]$ 即可得到条件二。

定理2中的条件二对于所有码本都成立, 主要通过假设 $K_a - 1$ 个活跃用户的信道衰落系数和发送的信息已知时, 将多用户问题退化到单用户问题, 再利用单用户情况下列表译码的理论界分析得出。但是, 当 K_a 很大时, 多用户干扰会变得非常严重, 已知 $K_a - 1$ 个活跃用户的信道衰落系数和发送信息的假设过于乐观, 因此该逆定理会变得非常松。为解决这个问题, 条件一直接处理多用户随机接入问题, 不依赖于上述假设, 但条件一仅针对高斯码本的情况成立。因此, 条件一在活跃用户较多时为主要约束条件, 条件二在活跃用户较少时为主要约束条件。

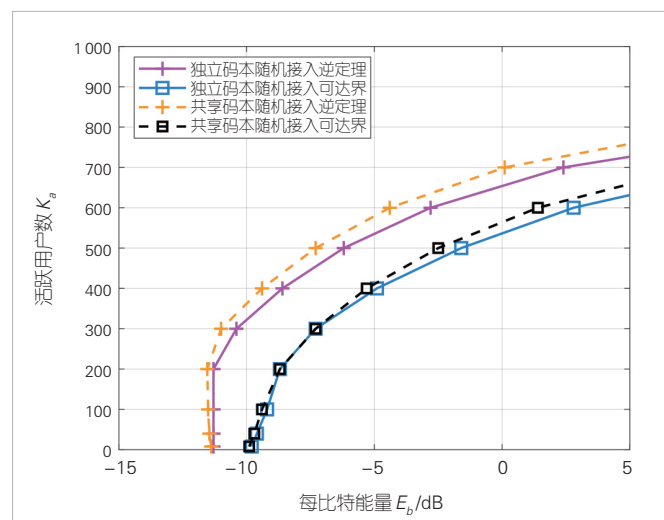
4 仿真结果分析

图1比较了在共享码本随机接入场景中, 采用不同每比特能量 E_b 时的平均每用户误差概率 P_e 。仿真中设置码长 $n = 1000$, 用户载荷 $J = 100$ bit, 基站天线数 $L \in \{32, 128\}$ 。图1中的可达界源于定理1, 逆定理源于定理2。从图中可以看出, 随着每比特能量的增加, 平均每用户误差概率逐渐减小。当基站天线数从 $L=32$ 增加到 $L=128$ 时, 达到某一误差概率要求所需的每比特能量显著减小, 因此基站配置多天线对支持大量用户可靠接入具有重要意义。此外, 当天线数较多且误差要求较高时, 可达界和逆定理之间的差距较小。具体而言, 当天线数 $L=128$ 且要求平均每用户误差概率 $P_e \leq 0.0001$ 时, 可达界和逆定理之间的差距能够减小到1 dB。

图2比较了本文推导的共享码本随机接入场景中的理论界以及文献[11]给出的信道状态信息未知独立码本随机接入时的理论界。仿真中假设活跃用户数 K_a 已知, 并设置码长 $n = 1000$, 用户载荷 $J = 100$ bit, 基站天线数 $L = 32$ 以及平均每用户误差概率要求 $P_e \leq 0.001$ 。对于独立码本随机接入, 设置 $K_a/K = 0.4$ 。图2还比较了共享码本随机接入和独立码本随机接入两种场景中可支持活跃用户数和每比特能量的理论关系。仿真结果表明, 当活跃用户数小于400时, 两种场景中的可达界和逆定理之间的差距均小于4 dB。两种场景中均存在多用户干扰消除现象, 即当活跃用户数小于某个门限时, 几乎不需要额外增加每比特能量即可满足误差要求。当活跃用户数大于200且进一步增加时, 所需的每比特能量也



▲图1 共享码本随机接入采用不同每比特能量的平均每用户误差概率



▲图2 独立码本以及共享码本随机接入采用不同每比特能量时可支持活跃用户数对比

随之增加。此外,在给定用户载荷、码长、基站天线数和最大允许误差的情况下,与独立码本随机接入相比, K_a 个活跃用户共享码本随机接入时所需的每比特能量较少。这是因为当用户共享同一个码本时,在译码过程中只需搜索一个码本空间即可,因而支持 K_a 个活跃用户可靠接入所需的每比特能量减少,且用户数越多,共享码本随机接入节省的每比特能量越多。

5 结束语

面向未来 umMTC 场景中的巨连接、零星活跃、小包传输和低功耗的需求,本文围绕共享码本随机接入信息理论极限分析的关键科学问题,建立了非渐近域下大规模随机接入系统传输性能紧致的理论极限界,系统性地分析天线数、码长、用户载荷、发送功率、平均每用户误差概率等参数之间的非渐近理论关系。仿真结果证实了推导的非渐近理论界的紧致性,并且表明当活跃用户数量低于某一门限时,存在近似多用户干扰消除的效果。此外,通过在基站配置大量天线,可以有效降低平均每用户误差概率,从而保证大量用户能够可靠地接入通信系统。与独立码本随机接入相比,用户共享同一码本时数据译码所需搜索的空间较小,因此共享码本随机接入能够实现更高的能量效率。本文建立的共享码本随机接入非渐近理论界对探索 umMTC 场景的性能极限具有重要意义。

参考文献

- [1] POLYANSKIY Y. A perspective on massive random-access [C]// Proceedings of IEEE International Symposium on Information Theory (ISIT). IEEE, 2017: 2523–2527. DOI: 10.1109/ISIT.2017.8006984
- [2] KOWSHIK S S, ANDREEV K, FROLOV A, et al. Energy efficient coded random access for the wireless uplink [J]. IEEE transactions on communications, 2020, 68(8): 4694–4708. DOI: 10.1109/TCOMM.2020.3000635
- [3] KOWSHIK S S, ANDREEV K, FROLOV A, et al. Energy efficient random access for the quasi-static fading MAC [C]// Proceedings of IEEE International Symposium on Information Theory (ISIT). IEEE, 2019: 2768–2772. DOI: 10.1109/ISIT.2019.8849288
- [4] NGO K H, LANCHO A, DURISI G, et al. Massive uncoordinated access with random user activity [C]// Proceedings of IEEE International Symposium on Information Theory (ISIT). IEEE, 2021: 3014–3019. DOI: 10.1109/ISIT45174.2021.9518253
- [5] RAVI J, KOCH T. Scaling laws for many-access channels and unsourced random access [C]// Proceedings of 55th Asilomar Conference on Signals, Systems, and Computers. IEEE, 2021: 1482–1487. DOI: 10.1109/IEEECONF53345.2021.9723116
- [6] FENGLER A, HAGHIGHATSHOAR S, JUNG P, et al. Non-bayesian activity detection, large-scale fading coefficient

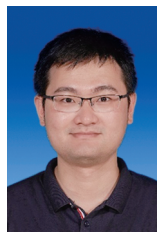
estimation, and unsourced random access with a massive MIMO receiver [J]. IEEE transactions on information theory, 2021, 67(5): 2925–2951. DOI: 10.1109/TIT.2021.3065291

- [7] POLYANSKIY Y, POOR H V, VERDU S. Channel coding rate in the finite blocklength regime [J]. IEEE transactions on information theory, 2010, 56(5): 2307–2359. DOI: 10.1109/TIT.2010.2043769
- [8] FANO R M. Transmission of information [M]. New York: John Wiley and Sons, 1961
- [9] PANG L H, ZHANG J, ZHANG Y, et al. Investigation and comparison of 5G channel models: from QuaDRiGa, NYUSIM, and MG5G perspectives [J]. Chinese journal of electronics, 2022, 31(1): 1–17. DOI: 10.1049/cje.2021.00.103
- [10] REEVES G, GASTPAR M C. Approximate sparsity pattern recovery: information-theoretic lower bounds [J]. IEEE transactions on information theory, 2013, 59(6): 3451–3465. DOI: 10.1109/TIT.2013.2253852
- [11] GAO J Y, WU Y P, SHAO S, et al. Energy efficiency of massive random access in MIMO quasi-static Rayleigh fading channels with finite blocklength [J]. IEEE transactions on information theory, 2023, 69(3): 1618–1657. DOI: 10.1109/TIT.2022.3220261

作者简介



高俊园, 上海交通大学电子信息与电气工程学院在读博士研究生; 主要研究方向为大规模随机接入、有限块长信息理论分析等。



吴泳澎, 上海交通大学教授, IEEE Senior Member, IEEE ICC、Globecom、VTC 等国际著名会议技术委员会主席及委员; 研究方向为无线通信理论与关键技术; 主持国家自然科学基金、国家重点研发计划子课题, 负责中兴通讯、国家电网等的多个科研项目; 获国家自然科学基金委优秀青年基金、IEEE 通信学会亚太区杰出青年研究学者奖、季寒冰青年教师奖、中国科协“青年托举人才计划”和 2014 德国洪堡学者等荣誉, 博士论文《多用户 MIMO 传输理论方法》获首届中国通信学会“全国信息通信领域优秀博士学位论文”奖, 并获多个学术会议最佳论文奖; 发表论文 90 余篇。



张文军, 上海交通大学教授、数字电视国家工程研究中心首席科学家、未来媒体网络协同创新中心主任; 主要研究方向为视频编码与无线传输、多媒体语义分析与广播/宽带网络融合; 是 2006 年发布的中国 DTTB 标准 (DTMB) 的主要贡献者; 发表论文 130 余篇, 拥有专利 238 项。

大规模离散 MU-MIMO: 低复杂度、信息理论最优检测与 多用户编码



Massive Discrete MU-MIMO: Low-Complexity Information-Theoretically Optimal Detection and Multi-User Coding

陈学辉/CHEN Xuehui¹, 池育浩/CHI Yuhao¹, 刘雷/LIU Lei²

(1. 西安电子科技大学, 中国 西安 710071;

2. 浙江大学, 中国 杭州 310058)

(1. Xidian University, Xi'an 710071, China;

2. Zhejiang University, Hangzhou 310058, China)

DOI: 10.12142/ZTETJ.202401009

网络出版地址: <http://kns.cnki.net/kcms/detail/34.1228.TN.20240228.0939.004.html>

网络出版日期: 2024-02-28

收稿日期: 2024-01-10

摘要: 研究了一种实际的大规模离散多用户多输入多输出 (MU-MIMO) 系统, 涵盖了大规模天线与用户、实用的信道编译码、任意输入分布、仅接收机已知信道状态信息、一般右酉不变信道矩阵、模数转化器等实际通信约束。针对理想和低分辨率模数转换器下的大规模离散 MU-MIMO 系统, 分别提出了低复杂度、信息理论最优接收机和多用户码设计准则, 并给出实际多用户低密度奇偶校验码 (LDPC) 设计。数值实验证实所设计的多用户 LDPC 比现有的单用户 LDPC 获得了 2~5 dB 性能增益, 同时还解决在信道恶劣和低分辨率模数转换器下单用户码无法准确恢复信息的难题, 揭示了现有贝叶斯最优接收机与单用户码直接结合不再是最优方案。

关键词: 大规模离散 MU-MIMO; 低复杂度; 信息理论最优; 迭代接收机; 多用户码

Abstract: A practical massive discrete multi-user multiple-input multiple-output (MU-MIMO) system is studied, which includes massive antennas and users, practical channel coding and decoding, arbitrary input distributions, available channel state information at the receiver, general right-unitarily-invariant channel matrices, and the analog-to-digital converter. For massive discrete MU-MIMO systems with ideal and low-resolution analog-to-digital converters, the low-complexity and information-theoretically optimal receiver and multi-user code design principles are proposed, respectively, on which practical multi-user low-density parity-check codes (LDPC) are designed. Numerical results show that the designed multi-user LDPC can achieve a performance gain of up to 2~5 dB over the existing single-user LDPC. They also address the issue that single-user codes cannot successfully recover information under poor channel and low-resolution analog-to-digital converters, and reveal that the direct combinations of existing Bayes-optimal receivers with single-user codes are no longer optimal.

Keywords: massive discrete MU-MIMO; low complexity; information-theoretically optimal; iterative receiver; multi-user code

引用格式: 陈学辉, 池育浩, 刘雷. 大规模离散 MU-MIMO: 低复杂度、信息理论最优检测与多用户编码 [J]. 中兴通讯技术, 2024, 30(1): 46-54. DOI: 10.12142/ZTETJ.202401009

Citation: CHEN X H, CHI Y H, LIU L. Massive discrete MU-MIMO: low-complexity information-theoretically optimal detection and multi-user coding [J]. ZTE technology journal, 2024, 30(1): 46-54. DOI: 10.12142/ZTETJ.202401009

随着无线通信技术和物联网行业的快速发展, 一系列新兴通信服务不断涌现, 包括异构车联网系统、多媒体通讯以及移动无人机网络等。预计在未来 6G 无线通信场景

中, 每平方公里的无线终端连接数量可能高达千万级别^[1-2]。面对如此大规模连接和高吞吐量的要求, 多址接入技术是一项关键的解决方案。然而, 传统的正交多址接入 (OMA) 技术在有限的频谱资源上仅能支持少量用户。相比之下, 非正交多址接入 (NOMA) 允许大量用户共享频谱传输, 从而提升接入设备数量、频谱利用率和降低通信时延。为了进一步提升空间资源利用率, NOMA 技术与多用户多输入多输出

基金项目: 国家自然科学基金重大项目 (62394292); 国家自然科学基金重点项目 (62131016); 国际自然科学基金青年项目 (62201424、62301485); 中兴产学研项目 (IA20231213009); 国家重点研发计划 (2021YFA1000500); 陕西省重点研发计划 (2023-YBGY-218)

(MU-MIMO)^[3]技术相结合。然而,由于未来6G业务数据种类繁多、通信场景多样化、用户和天线数量庞大,通信场景变得更加复杂,这导致现有基于理想假设的MU-MIMO技术不再适用,无法有效支持复杂的6G通信场景。另外,现有工作只考虑部分非理想假设下的MU-MIMO。例如,文献[4]只考虑1 bit量化约束下的MU-MIMO系统的预编码。文献[5]只考虑贝叶斯最优的接收机设计,忽略了信道编码的影响。文献[6]只考虑点对点信道下的码设计,忽略了用户间干扰的影响。总而言之,现有技术很少综合考虑多种非理想假设下的MU-MIMO接收机和编码设计。因此,本文聚焦于更实用的大规模离散MU-MIMO^[7],考虑了更为实际的系统假设:包括大规模天线与用户、实用的信道编译码、任意输入分布、仅接收机已知信道状态信息(CSI)、一般右酉不变信道矩阵(涵盖常见的瑞利、相关、病态等信道矩阵)、模数转换器(ADC)等实际通信约束。这些实际的假设也给大规模离散MU-MIMO系统设计带来了新的挑战:

1) 如何设计适合于大规模离散MU-MIMO的低复杂度、信息理论最优接收机是一个巨大挑战。针对更为实际的通信假设,现有的接收机无法达到最优性能,因此,迫切需要设计低复杂度、高可靠的接收机。

2) 大规模离散MU-MIMO系统的实际多用户码设计原理尚不明确。传统的单用户信道编码仅仅只考虑了信道噪声的影响,而忽略了大规模天线和多用户间干扰。因此,现有的信道编码方案无法直接应用在实际大规模离散MU-MIMO系统中,迫切需要针对复杂的实际通信场景设计更加灵活和实用的多用户编码方案。

为解决上述难题,本文提出了低复杂度、信息理论最优接收机和多用户码设计准则。具体而言,首先在理想ADC器件的假设下,利用互信息与最小均方误差定理^[8],提出了

容量-面积定理,证实近似消息传递(AMP)、正交AMP和记忆AMP的容量最优性,并给出最优多用户码设计准则。进一步考虑实际低分辨率ADC器件影响,对广义AMP、广义正交AMP和广义记忆AMP接收机给出了可达速率分析以及相应的最优码设计准则。最后,本文以正交AMP、记忆AMP以及广义正交AMP接收机为例,设计了多用户低密度奇偶校验码(MU-LDPC)。数值实验表明,所设计的MU-LDPC比现有的单用户LDPC(SU-LDPC)在误码率(BER)性能上提高2~5 dB,特别是在信道条件恶劣和1 bit量化场景下,现有的规则和非规则的SU-LDPC已经无法准确恢复信息,而MU-LDPC仍能够准确恢复信息。这为未来6G复杂场景通信提供了有效的解决方案。

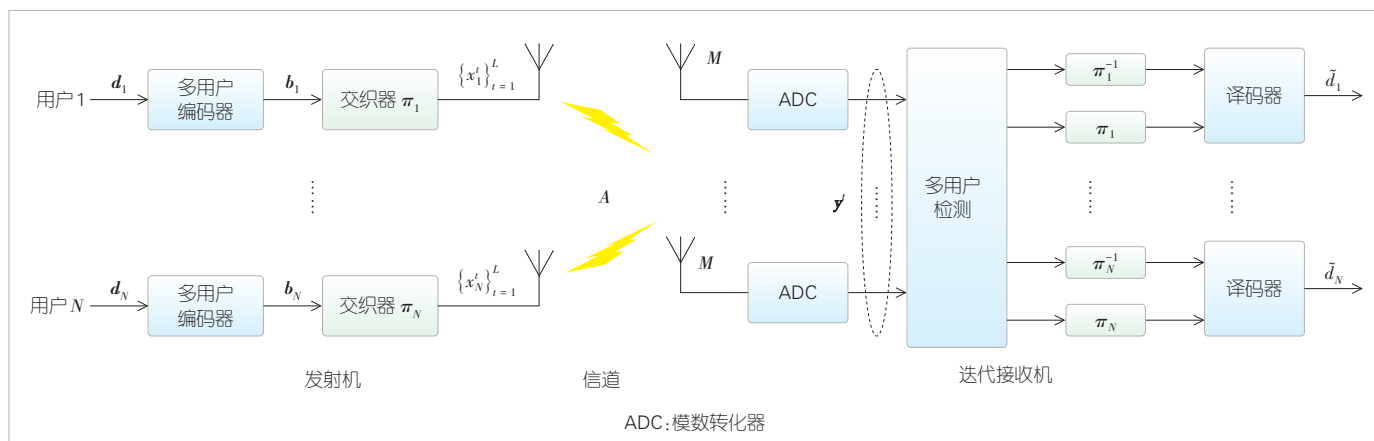
1 系统模型

如图1所示,在大规模离散MU-MIMO系统中, N 个单天线的用户采用交织多址接入方式进行传输,每个用户的传输符号长度为 L 。具体而言,用户 k 采用多用户编码器(包含信道编码和离散数字调制)对信息序列 $\mathbf{d}_k$ 进行编码得到编码序列 $\mathbf{b}_k$,再通过每个用户特定的交织器 π_k 生成符号序列 $\mathbf{x}_k$,并将其发送至信道,其中 $k = 1, 2, \dots, N$ 。为简化表示,在 t 时刻,将所有用户的发送信号记为 $\mathbf{x}^t = [x_1^t, x_2^t, \dots, x_N^t]^T$ 。

假设接收机配备了 M 根接收天线,考虑到实际ADC对接收到的信号进行量化处理,我们将 t 时刻量化后的信号记为 $\mathbf{y}^t$,即:

$$\mathbf{y}^t = Q(\mathbf{A}\mathbf{x}^t + \mathbf{n}^t), \quad t = 1, 2, \dots, L, \quad (1)$$

其中 $\mathbf{A}$ 为右酉不变信道矩阵(涵盖常见的瑞利、相关、病态等信道矩阵), $\mathbf{n}^t$ 为加性高斯白噪声, $Q(\cdot)$ 为逐符号量化函数,即对复信号的实部和虚部分别进行量化。基站利用接收信号 $\mathbf{y}^t$ 、信道矩阵 $\mathbf{A}$ 、量化函数 $Q(\cdot)$ 和 $\mathbf{x}^t$ 的先验信息,采用



▲图1 大规模离散多用户多输入多输出系统

迭代接收机实现信号恢复。

本文研究的大规模离散 MU-MIMO 系统基于以下假设：

1) 系统中包含大规模的用户和天线，即 $N, M \rightarrow \infty$ 且固定信道负载 $\delta = N/M$ 。

2) 信道矩阵 $\mathbf{A}$ 的奇异值分解为 $\mathbf{A} = \mathbf{U}\mathbf{\Lambda}\mathbf{V}^H$ ，其中 $\mathbf{U} \in \mathbb{C}^{M \times M}$ 和 $\mathbf{V} \in \mathbb{C}^{N \times N}$ 为酉矩阵， $\mathbf{\Lambda} \in \mathbb{C}^{M \times N}$ 为对角矩阵。 $\mathbf{U}\mathbf{\Lambda}$ 和 $\mathbf{V}$ 是独立的，而且 $\mathbf{V}$ 服从 Haar 分布。

3) 仅接收端已知信道矩阵 $\mathbf{A}$ ，而发送端未知。

2 低复杂度、容量最优接收机和多用户码设计

由于接收机是逐符号检测的，为简化表示，在本文余下部分忽略上标 t 。在理想 ADC 器件条件下（即不考虑非线性预处理的影响），公式 (1) 退化为标准线性模型：

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n}. \quad (2)$$

针对该模型，本文采用能够有效克服噪声和用户间干扰的 AMP 类接收机恢复信号。然而，现有的接收机设计常常忽略信道编码约束的影响。通常，接收机的设计目标是达到 MMSE 性能最优，但无法保证无差错信号恢复。因此，本文将在实际通信约束下研究低复杂度、容量最优接收机和多用户码设计准则。

2.1 AMP 类接收机

如图 1 所示，迭代接收机是由多用户检测器和一组译码器组成，其一般形式表示为：

$$\text{检测器: } \mathbf{r}_t = \gamma_t(\mathbf{x}_t), \quad (3)$$

$$\text{译码器: } \mathbf{x}_{t+1} = \phi_t(\mathbf{r}_t). \quad (4)$$

AMP 类接收机为迭代接收机特例，具体如下：

1) AMP^[9]接收机是由匹配滤波器 (MF) 和后验概率译码器组成，即：

$$\mathbf{r}_t = \gamma_t(\mathbf{x}_t) = \mathbf{x}_t + \mathbf{A}^H(\mathbf{y} - \mathbf{A}\mathbf{x}_t) + \mathbf{r}_t^{\text{Onsager}}, \quad (5)$$

$$\mathbf{x}_{t+1} = \phi_t(\mathbf{r}_t) = \mathbb{E}\{\mathbf{x}|\mathbf{r}_t\}, \quad (6)$$

其中 $\mathbf{r}_t^{\text{Onsager}} = \delta < \phi'_t(\mathbf{r}_{t-1}) > (\mathbf{r}_{t-1} - \mathbf{x}_{t-1})$ ， $\phi'_t(\cdot)$ 表示 $\phi_t(\cdot)$ 的导函数， $< \cdot >$ 表示取平均。

2) 正交 AMP^{[10][11]}接收机由正交线性最小均方误差 (LMMSE) 检测器和正交后验概率译码器组成，即：

$$\mathbf{r}_t = \gamma_t(\mathbf{x}_t) = \mathbf{x}_t + \frac{1}{\varepsilon_t^\gamma} \mathbf{A}^H(\rho_t \mathbf{I} + \mathbf{A}\mathbf{A}^H)^{-1}(\mathbf{y} - \mathbf{A}\mathbf{x}_t), \quad (7)$$

$$\mathbf{x}_{t+1} = \phi_t(\mathbf{r}_t) = \frac{1}{\varepsilon_{t+1}^\phi} [\hat{\phi}_t(\mathbf{r}_t) + (1 - \varepsilon_{t+1}^\phi)\mathbf{r}_t], \quad (8)$$

其中 $\rho_t = \sigma^2/v_t^x$ ， σ^2 表示噪声方差， v_t^x 表示检测器 $\gamma_t(\mathbf{x}_t)$ 的输出方差， $\hat{\phi}_t(\mathbf{r}_t) = \mathbb{E}\{\mathbf{x}|\mathbf{r}_t\}$ 。 ε_t^γ 和 ε_{t+1}^ϕ 为正交系数，以保证输出与输入误差的正交性，具体表达式为 $\varepsilon_t^\gamma = (1/M)\text{tr}\{\mathbf{A}^H(\rho_t \mathbf{I} + \mathbf{A}\mathbf{A}^H)^{-1}\mathbf{A}\}$ ， $\varepsilon_{t+1}^\phi = 1 - (1/(Nv_t^x))\|\mathbb{E}\{\mathbf{x}|\mathbf{x} + \sqrt{v_t^x}\boldsymbol{\eta}\} - \mathbf{x}\|^2$ ， $\boldsymbol{\eta} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$ 。

3) 记忆 AMP^[12]接收机是由正交记忆 MF、正交后验概率译码器以及 damping 组成，即：

$$\mathbf{r}_t = \gamma_t(\mathbf{x}_t) = \frac{1}{\varepsilon_t} \left(\mathbf{A}^H \hat{\mathbf{r}}_t + \sum_{i=1}^l p_{t,i} \mathbf{x}_i \right), \quad (9)$$

$$\mathbf{x}_{t+1} = [\mathbf{x}_1, \dots, \mathbf{x}_t, \phi_t(\mathbf{r}_t)] \boldsymbol{\zeta}_{t+1}, \quad (10)$$

其中 $\phi_t(\mathbf{r}_t)$ 与正交 AMP 接收机的正交后验概率译码函数一致， $\hat{\mathbf{r}}_t = \theta_t \mathbf{B} \hat{\mathbf{r}}_{t-1} + \xi_t(\mathbf{y} - \mathbf{A}\mathbf{x}_t)$ ， $\mathbf{B} = \lambda^\dagger \mathbf{I} - \mathbf{A}\mathbf{A}^H$ ， $\lambda^\dagger = (\lambda_{\min} + \lambda_{\max})/2$ 表示 $\mathbf{A}\mathbf{A}^H$ 的最小特征值 $\lambda_{\min}$ 和最大特征值 $\lambda_{\max}$ 的平均值， ε_t 和 $p_{t,i}$ 为正交系数，以保证输出与所有输入误差的正交性。 $\boldsymbol{\zeta}_t = [\zeta_{t,1}, \zeta_{t,2}, \dots, \zeta_{t,l_t}]^T$ 表示 damping 向量， $l_t = \min\{L, t\}$ 为 damping 长度， L 为最大 damping 长度。

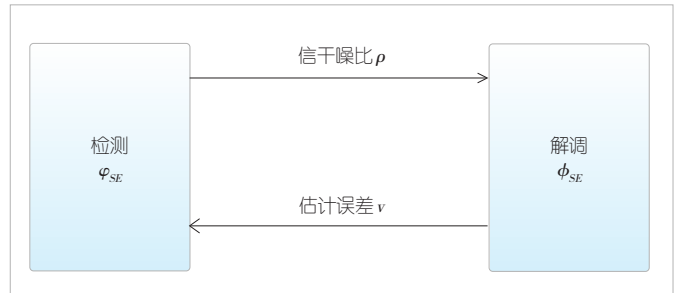
2.2 容量-面积定理

针对任意离散输入信号 $\mathbf{x}$ 和右酉不变信道矩阵 $\mathbf{A}$ ，离散 MU-MIMO 的约束容量为：

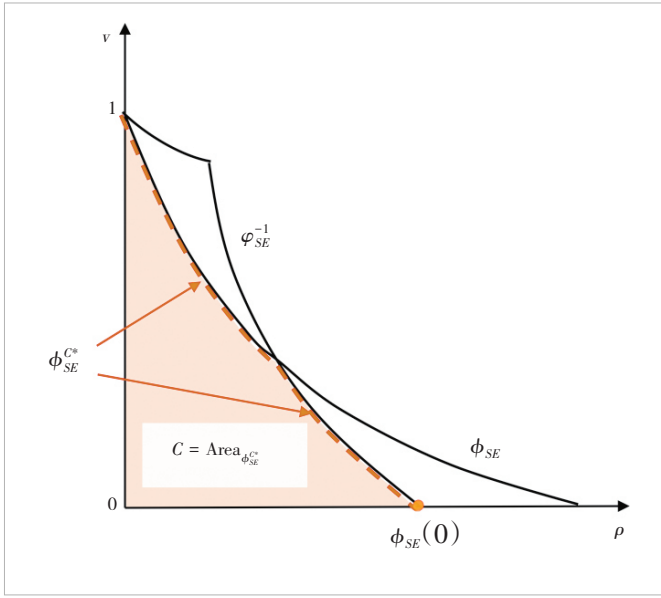
$$C = I(\mathbf{A}\mathbf{x} + \mathbf{n}; \mathbf{x}). \quad (11)$$

然而，对于大规模系统，直接计算公式 (11) 中的互信息来表征约束容量会非常困难。如图 2 所示，借助 AMP 类算法的 MMSE 最优性、互信息与最小均方误差 (I-MMSE) 定理^[7]和状态演化 (SE)，简化大规模离散 MU-MIMO 的约束容量的计算，并证明 AMP 类接收机的容量最优性。

定理 1 (容量-面积定理)：如图 3 所示，假设不动点唯一，针对大规模离散 MU-MIMO 场景，给定信噪比下，根据



▲图2 近似消息传递(AMP)类算法的状态演进



▲图3 大规模离散多用户多输入多输出容量-面积定理

I-MMSE 定理^[8], 由 AMP 类接收机中检测器和解调器的 SE 可计算大规模离散 MU-MIMO 的约束容量为:

$$C = \int_0^{\varphi_{SE}(0)} \phi_{SE}^C(\rho) d\rho, \quad (12)$$

其中 $\phi_{SE}^C(\rho) = \min\{\phi_{SE}(\rho), \varphi_{SE}^{-1}(\rho)\}$, $\phi_{SE}(\cdot)$ 表示解调器的 SE 函数, $\varphi_{SE}(\cdot)$ 表示检测器的 SE 函数, $\varphi_{SE}^{-1}(\cdot)$ 表示 $\varphi_{SE}(\cdot)$ 的反函数. $\phi_{SE}(\cdot)$ 与 $\varphi_{SE}(\cdot)$ 具体定义如下:

$$\phi_{SE}(\rho) = \frac{1}{N} E\{\|\phi(\mathbf{x} + \sqrt{\rho^{-1}} \mathbf{n}) - \mathbf{x}\|^2\}, \quad (13)$$

$$\varphi_{SE}(\rho) = \frac{1}{N} E\{\|\varphi(\mathbf{x} + \sqrt{\rho^{-1}} \xi) - \mathbf{x}\|^2\}, \quad (14)$$

其中 $\mathbf{n} \sim \mathcal{CN}(0, \mathbf{I})$, $\xi \sim \mathcal{CN}(0, \mathbf{I})$, $\phi(\mathbf{r}) = E\{\mathbf{x}|\mathbf{r}\}$ 表示解调器的估计函数, $\varphi(\cdot)$ 表示检测器函数 $\gamma(\cdot)$ 与可能存在的正交, damping 等部分组成的检测函数, 具体表达式由接收机结构确定。

由于信道编码获得了编码增益, 译码器能够输出比解调器更低的均方误差 (MSE), 即 $\phi_{SE}^C(\rho) < \phi_{SE}(\rho)$ 。同时, 根据无差错译码准则, 需要保证译码器的 SE 与检测器的 SE 之间有足够的译码通道, 即 $\phi_{SE}^C(\rho) < \varphi_{SE}^{-1}(\rho)$ 。当译码器的 SE 函数 $\phi_{SE}^C(\rho)$ 固定时, 根据 I-MMSE 定理, AMP 类接收机的可达速率为:

$$R_{AMP \text{ 类接收机}} = \int_0^{\varphi_{SE}(0)} \phi_{SE}^C(\rho) d\rho, \quad \text{其中 } \phi_{SE}^C(\rho) <$$

$\min\{\phi_{SE}(\rho), \varphi_{SE}^{-1}(\rho)\}$, $0 \leq \rho \leq \varphi_{SE}(0)$ 。基于以上条件, 可得 AMP 类接收机的约束容量最优性定理。

定理 2 (AMP 类接收机的约束容量最优性): 在 AMP 类接收机中, 假设不动点唯一, 根据译码器的输出 MSE 低于解调器、无差错译码的条件以及 I-MMSE 定理, 可得检测器和解调器的 SE 围成的面积为接收机的信息理论极限 (最大可达速率), 即:

$$R_{AMP \text{ 类接收机}}^{\max} \rightarrow \int_0^{\varphi_{SE}(0)} \phi_{SE}^C(\rho) d\rho, \quad (15)$$

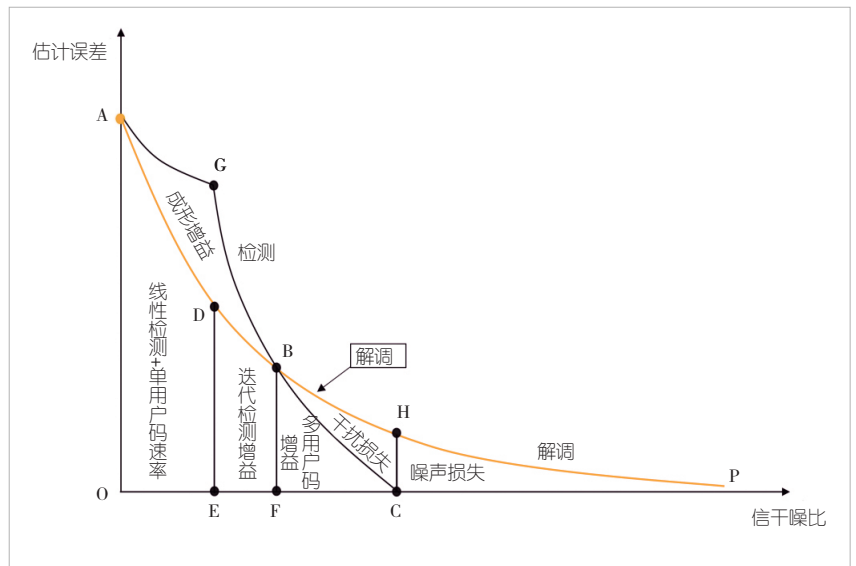
其中 $\phi_{SE}^C(\rho) = \min\{\phi_{SE}(\rho), \varphi_{SE}^{-1}(\rho)\}$ 。由定理 1 可知, AMP 类接收机可以达到大规模离散 MU-MIMO 的约束容量, 即证实 AMP 类接收机的约束容量最优性。

如图 4 所示, 给定接收机的结构, 根据不同约束条件下, 基于 I-MMSE 定理, 利用检测器、解调器和译码器的 SE, 可得如下面积定理。具体的推导过程与定理二类似, 这里不再详细展开。

1) AHPO 面积代表星座熵, 表示无用户间干扰、无噪声情况下接收机的最大可达速率。

2) AGB 面积代表高斯信号的成形增益。曲线 AG 表示未编码高斯信号的解调曲线, 是未编码离散信号解调曲线的上界。AGCO 面积表示高斯信号下的约束容量, ABCO 面积表示离散信号下的约束容量, 因此相差的 AGB 面积代表高斯信号的成形增益。

3) BHC 面积代表用户间干扰造成的容量损失。AHCO 面积表示点对点 AWGN 信道下的约束容量, ABCO 面积表示存在用户间干扰情况下的约束容量, 因此相差的 BHC 面积



▲图4 多用户多输入多输出面积定理示意图

代表干扰造成的容量损失。

4) HPC 面积代表信道噪声造成的容量损失。随着信噪比趋于无穷大, C 点会向右移动到无穷大(例如 P 点), AHPO 面积代表无噪情况下的约束容量, AHCO 面积代表确定信噪比下的约束容量, 因此相差的 HPC 面积代表噪声造成的容量损失。

5) ADEO 面积代表线性检测器和单用户译码器串行级联接收机的可达速率。如图 5 (a) 所示, 该接收机的检测器、解调器与译码器之间没有迭代。线性检测后到达 G 点, 解调后至 D 点, 译码后到达 E 点。

6) ABFO 面积代表 AMP 类迭代接收机和单用户译码器串行级联接收机的可达速率。如图 5 (b) 所示, 该接收机的检测器与解调器进行充分迭代后到达 B 点, 并将估计信号输入至译码器, 译码后到达 F 点, 导致整个算法并不是容量最优的。

7) ABCO 面积代表多用户码的约束容量 C 和 AMP 类接收机的最大可达速率 R 。如图 5 (c) 所示, 该接收机联合解调与译码, 形成多用户译码与检测器进行充分迭代后到达 C 点, 此时整个算法为约束容量最优。

8) DBFE 面积为 ABFO 面积与 ADEO 面积差, 代表对比线性检测接收机, 迭代接收机的可达速率增益。

9) BCF 面积为 ABCO 面积与 ABFO 面积差, 代表对比单用户码, 多用户码的可达速率增益。

2.3 最优码设计准则

为确保 AMP 类接收机达到最大传输速率, 需要根据 AMP 类接收机的迭代消息更新规则, 结合相应的 SE 分析, 来确定相应的最优多用户编码。

定理 3 (最优多用户码设计准则): 如图 3 所示, 以最大化 AMP 类接收机的传输速率为目标, 最优多用户译码器的 SE 为:

$$\phi_{SE}^c(\rho) \rightarrow \min \{ \phi_{SE}(\rho), \varphi_{SE}^{-1}(\rho) \}。 \quad (16)$$

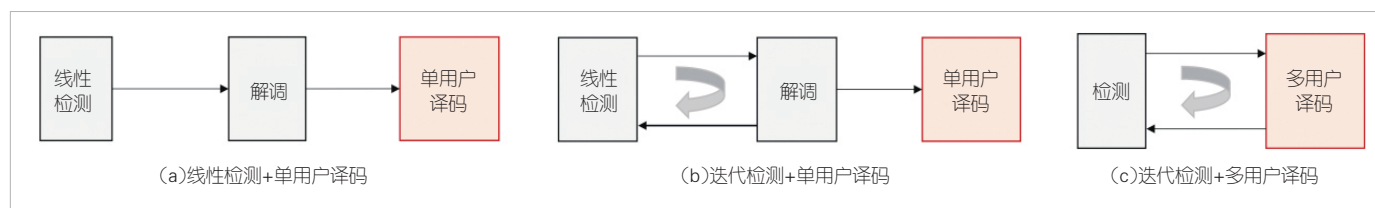
表 1 总结了常见迭代接收机, 其中 T 为接收机最大迭代次数。

2.4 数值实验

本节的仿真为基带的链路仿真, 用户数 $M = 500$, 服从均匀分布, 发送功率为 1。发送端采用 MU-LDPC, 接收端采用正交 AMP/记忆 AMP/Turbo-LMMSE 接收机进行信号检测, 信道为一般右西不变信道, 具体模型为: 信道矩阵 A 的奇异值分解为 $A = U\Lambda V^H$, 将 Λ 中的特征值 $\{d_i\}$ 设置为: $d_i/d_{i+1} = \kappa^{1/T}, i = 1, 2, \dots, T, \sum_{i=1}^T d_i^2 = N$, 其中 $T = \min \{M, N\}$, $\kappa > 1$ 表示 A 的条件数, U 和 V 由独立同分布高斯矩阵的 QR 分解生成。信道矩阵 A 仅接收端已知, 发送端未知。值得说明的是, 一般右西不变信道适用性更广泛, 涵盖了常见的瑞利、相关、病态以及部分大尺度模型的信道矩阵。在一般右西不变信道下工作的 AMP 类接收机, 在一些特定的信道中往往也能够工作。

图 6 给出了以高斯信号为例, 比较 AMP 接收机采用多用户码和单用户码方案的可达速率。可以看出, 随着信噪比的提升, 采用单用户码的 AMP 接收机可达速率趋于常值, 而采用多用户码的 AMP 接收机可达速率不断增加。在信噪比为 20 dB 时, 采用多用户码的方案比单用户码的方案获得了大约 2 倍的可达速率增益。

图 7 给出了四相相移键控 (QPSK) 调制, 当信道矩阵 A 的条件数 $\kappa = 50$ 、接收天线数 $M = 333$ 时, 针对记忆 AMP、

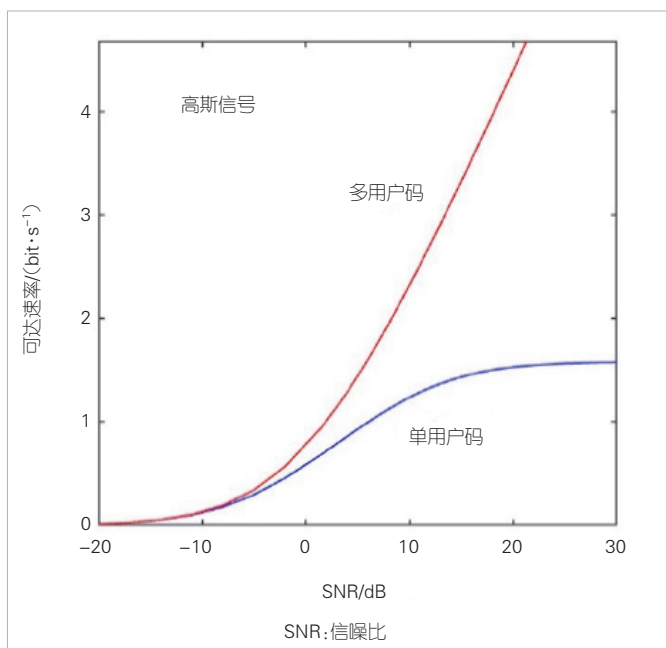


▲图 5 不同检测方案示意图

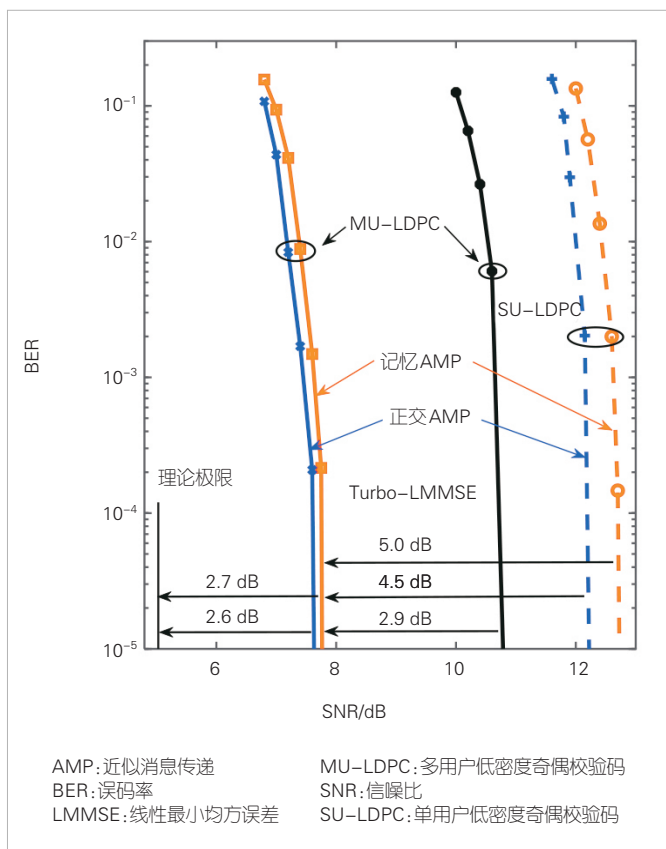
▼表 1 MU-MIMO 系统下 AMP/Turbo 类接收机对比

方案/算法	信号	信道矩阵	容量类型	最优性	计算复杂度
AMP ^[9]	任意	独立同分布	约束容量	容量最优	低复杂度
记忆AMP ^[12]					$O(MNT)$
正交AMP ^[10-11]		右酉不变			高复杂度
Turbo-LMMSE ^[13]	高斯		高斯容量		$O(M^2NT+MNT)$

AMP: 近似消息传递 LMMSE: 线性最小均方误差 MU-MIMO: 多用户多输入多输出



▲图6 不同方案的可达速率比较



▲图7 正交AMP/记忆AMP/Turbo-LMMSE接收机BER比较

正交AMP和Turbo-LMMSE接收机，所设计的码率为0.5的MU-LDPC和码率为0.5的SU-LDPC的BER性能对比：

1) 针对MU-LDPC：记忆AMP与正交AMP的BER性能

相近，二者仅差0.1 dB，但记忆AMP的复杂度约为正交AMP的 $1/M$ ；

2) 针对记忆AMP和正交AMP：MU-LDPC比SU-LDPC获得了4.5~5 dB增益，证实了贝叶斯最优检测接收机与现有最优单用户码相结合并非是最优方案；

3) 接收机性能比较：记忆AMP和正交AMP的MU-LDPC均比Turbo-LMMSE的MU-LDPC获得了约2.9 dB增益。

3 低分辨率ADC下低复杂度、信息理论最优接收机和多用户码设计

考虑实际ADC器件约束下的低复杂度、信息理论最优接收机和多用户码设计，本文采用广义AMP类接收机进行信号恢复。为简化讨论，公式（1）等价转换为如下广义线性模型（GLM）：

$$\begin{cases} \mathbf{y} = Q(\mathbf{z}, \mathbf{n}), \\ \mathbf{z} = \mathbf{A}\mathbf{x}, \\ \mathbf{x} \in C, x_i \sim P_X(x_i), \forall i. \end{cases} \quad (17)$$

基于上述模型，可采用能够有效克服噪声和用户间干扰的广义AMP类接收机进行信号恢复。

3.1 广义AMP类接收机

如图8所示，广义AMP类接收机包含线性检测和非线性检测（包括MMSE解量化以及MMSE解调和后验概率译码），可表示为：

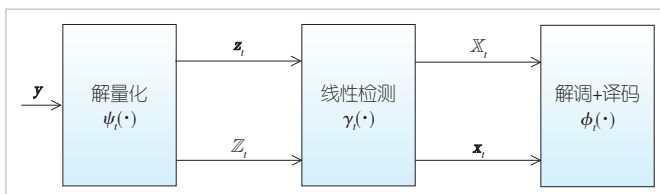
$$\text{线性检测: } \begin{bmatrix} \mathbf{x}_i \\ \mathbf{z}_i \end{bmatrix} = \begin{bmatrix} \phi_i(\mathbb{X}_i) \\ \psi_i(\mathbb{Z}_i) \end{bmatrix}, \quad (18)$$

$$\text{非线性检测: } \begin{bmatrix} \mathbb{X}_i \\ \mathbb{Z}_i \end{bmatrix} = \gamma_i(\mathbf{x}_i, \mathbf{z}_i). \quad (19)$$

根据信号检测和消息传递更新方式的不同，分为不同类型的接收机，具体如下：

1) 广义AMP^[14]是由解量化器、MF线性检测器和后验概率译码器组成，即：

$$\begin{bmatrix} \mathbf{x}_i \\ \mathbf{z}_i \end{bmatrix} = \begin{bmatrix} \phi_i(\mathbb{X}_i) \\ \psi_i(\mathbb{Z}_i) \end{bmatrix} = \begin{bmatrix} E\{\mathbf{x}_i|\mathbb{X}_i\} \\ \mathbf{z}_i\{\mathbf{z}_i|\mathbb{Z}_i\} \end{bmatrix}, \quad (20)$$



▲图8 广义AMP类接收机

$$\begin{bmatrix} \mathbb{X}_t \\ \mathbb{Z}_t \end{bmatrix} = \begin{bmatrix} \mathbf{x}_t + \frac{\nu_{t+1}^x}{\nu_t^x} \mathbf{A}^H (\mathbf{z}_t + \mathbb{Z}_t) \\ \mathbf{A} \mathbf{x}_t - \frac{\nu_{t+1}^z}{\nu_t^z} (\mathbf{z}_t + \mathbb{Z}_t) \end{bmatrix}, \quad (21)$$

其中 ν_t^x 和 ν_t^z 分别表示线性检测端 $\phi_t(\mathbb{X}_t)$ 和 $\psi_t(\mathbb{Z}_t)$ 的输出方差。

2) 广义正交 AMP^[15]是由正交 LMMSE 检测器、正交解量化器和正交后验概率译码器组成, 即:

$$\begin{bmatrix} \mathbf{x}_t \\ \mathbf{z}_t \end{bmatrix} = \begin{bmatrix} \phi_t(\mathbb{X}_t) \\ \psi_t(\mathbb{Z}_t) \end{bmatrix} = \begin{bmatrix} \frac{1}{\varepsilon_t^\phi} [\mathbf{E}\{\mathbf{x}|\mathbb{X}_t\} + (\varepsilon_t^\phi - 1)\mathbb{X}_t] \\ \frac{1}{\varepsilon_t^\psi} [\mathbf{E}\{\mathbf{z}|\mathbb{Z}_t\} + (\varepsilon_t^\psi - 1)\mathbb{Z}_t] \end{bmatrix}, \quad (22)$$

$$\begin{bmatrix} \mathbb{X}_t \\ \mathbb{Z}_t \end{bmatrix} = \gamma_t(\mathbf{x}_t, \mathbf{z}_t) = \begin{bmatrix} \gamma_t^x(\mathbf{x}_t, \mathbf{z}_t) \\ \gamma_t^z(\mathbf{x}_t, \mathbf{z}_t) \end{bmatrix} = \begin{bmatrix} \frac{1}{\delta \varepsilon_t^\gamma} \hat{\gamma}_t(\mathbf{x}_t, \mathbf{z}_t) + \mathbf{x}_t \\ \frac{1}{1 - \varepsilon_t^\gamma} [\mathbf{A}(\hat{\gamma}_t(\mathbf{x}_t, \mathbf{z}_t) + \mathbf{x}_t) - \varepsilon_t^\gamma \mathbf{z}_t] \end{bmatrix}, \quad (23)$$

其中 $\rho_t = v_t^z/v_t^x$, v_t^x 和 v_t^z 分别为 $\phi_t(\mathbb{X}_t)$ 和 $\psi_t(\mathbb{Z}_t)$ 的输出方差, $\hat{\gamma}_t(\mathbf{x}_t, \mathbf{z}_t) \equiv \mathbf{A}^H(\rho_t \mathbf{I} + \mathbf{A} \mathbf{A}^H)^{-1}(\mathbf{z}_t - \mathbf{A} \mathbf{x}_t)$, $\{\varepsilon_t^\phi, \varepsilon_t^\psi, \varepsilon_t^\gamma\}$ 表示正交系数, 以确保输出与输入误差的正交性, 具体表达式为 $\varepsilon_t^\phi = 1 - (1/(N \mathbb{V}_t^x)) \|\mathbf{E}\{\mathbf{x}|\mathbf{x} + \sqrt{\mathbb{V}_t^x} \boldsymbol{\eta}\} - \mathbf{x}\|^2$, $\varepsilon_t^\psi = 1 - (1/(M \mathbb{V}_t^z)) \|\mathbf{E}\{\mathbf{z}|\mathbf{z}_t\} - \mathbf{z}\|^2$, $\varepsilon_t^\gamma = (1/M) \text{tr}\{\mathbf{A}^H(\rho_t \mathbf{I} + \mathbf{A} \mathbf{A}^H)^{-1} \mathbf{A}\}$, $\mathbb{V}_t^x$ 和 $\mathbb{V}_t^z$ 为 $\gamma_t(\mathbf{x}_t, \mathbf{z}_t)$ 的输出方差, $\boldsymbol{\eta} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$ 。

3) 广义记忆 AMP^[16]是由正交记忆 MF、正交解量化器和正交后验概率译码器以及 damping 组成, 即:

$$\begin{bmatrix} \mathbf{x}_t \\ \mathbf{z}_t \end{bmatrix} = \begin{bmatrix} \zeta_{t,l_t} \phi_t(\mathbb{X}_t) + \sum_{i=1}^{l_t-1} \zeta_{t,i} \mathbf{x}_{t-l_t+i} \\ \alpha_{t,l_t} \psi_t(\mathbb{Z}_t) + \sum_{i=1}^{l_t-1} \alpha_{t,i} \mathbf{z}_{t-l_t+i} \end{bmatrix}, \quad (24)$$

$$\begin{bmatrix} \mathbb{X}_{t+1} \\ \mathbb{Z}_{t+1} \end{bmatrix} = \gamma_t(\mathbf{X}_t, \mathbf{Z}_t) = \begin{bmatrix} \beta_t^x (\delta^{-1} \hat{\mathbb{X}}_t + \sum_{i=1}^l p_{t,i} \mathbf{x}_i) \\ \beta_t^z [\mathbf{A}(\hat{\mathbb{X}}_t + \frac{\xi_t}{\theta_t} \mathbf{x}_t) - \sum_{i=1}^l p_{t,i} \mathbf{z}_i] \end{bmatrix}, \quad (25)$$

其中, $\phi_t(\mathbb{X}_t)$ 和 $\psi_t(\mathbb{Z}_t)$ 与广义正交 AMP 的正交后验概率译码和正交解量化函数一致。 $\hat{\mathbb{Z}}_t = \theta_t \mathbf{B} \hat{\mathbb{Z}}_{t-1} + \xi_t (\mathbf{z}_t - \mathbf{A} \mathbf{x}_t)$, $\mathbf{B} = \lambda^+ \mathbf{I} - \mathbf{A} \mathbf{A}^H$ 和 $\lambda^+ = (\lambda_{\min} + \lambda_{\max})/2$ 分别表示 $\mathbf{A} \mathbf{A}^H$ 的最小特征值 $\lambda_{\min}$ 和最大特征值 $\lambda_{\max}$ 的平均值, $\{\beta_t^x, \beta_t^z, p_{t,i}\}$ 为正交系数, 以确保输出与所有输入误差的正交性。 $\zeta_t = [\zeta_{t,1}, \zeta_{t,2}, \dots, \zeta_{t,l_t}]^T$ 和

$\boldsymbol{\alpha}_t = [\alpha_{t,1}, \alpha_{t,2}, \dots, \alpha_{t,l_t}]^T$ 分别表示解调器的 damping 向量和解量化器的 damping 向, $l_t = \min\{L, t\}$ 为 damping 长度, L 为最大 damping 长度。

3.2 最大可达速率分析

注意到图 8 中广义 AMP 类接收机具有多层结构, 其中线性检测是双输入双输出, 因此无法直接采用定理 2 中的可达速率分析方法^[8]。为解决这一问题, 本文提出一种内迭代广义 AMP 类接收机。如图 9 所示, 解量化和线性检测通过内迭代形成增强的多用户检测。具体而言, 对于给定的输入信号, 解量化和检测之间进行充分迭代, 然后将估计信号 $\mathbb{X}_t$ 传递到解调和译码。值得说明的是, 尽管内迭代广义 AMP 类接收机与广义 AMP 接收机结构不同, 但它们的 SE 收敛点是相同的。因此可以借助内迭代广义 AMP 接收机的 SE 分析原始的广义 AMP 接收机的可达速率, 即借助 I-MMSE 定理 (如图 3 所示) 来分析其最大可达速率:

$$R_{\text{广义 AMP 类接收机}}^{\max} \rightarrow \int_0^{\varphi_{\text{SE}}(0)} \phi_{\text{SE}}^c(\rho) d\rho, \quad (26)$$

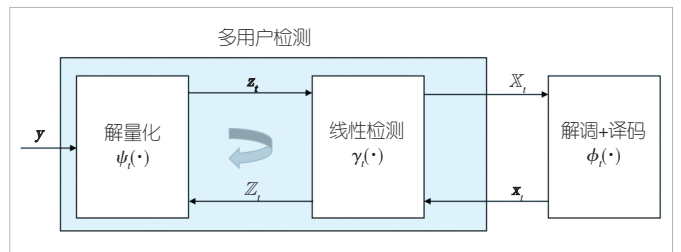
其中 $\phi_{\text{SE}}^c(\rho) = \min\{\phi_{\text{SE}}(\rho), \varphi_{\text{SE}}^{-1}(\rho)\}$, $\phi_{\text{SE}}(\cdot)$ 表示解调的 SE 函数, $\varphi_{\text{SE}}(\cdot)$ 表示增强多用户检测的 SE 函数, $\varphi_{\text{SE}}^{-1}(\cdot)$ 为 $\varphi_{\text{SE}}(\cdot)$ 的反函数。

由于具有任意离散输入信号和酉不变信道矩阵的 GLM 的约束容量仍未得到严格证明, 因此无法证明广义 AMP 类接收机的容量最优性。然而, 基于广义 AMP 类接收机在未编码 GLM 中的贝叶斯最优性, 我们推测其在最优编码设计准则下是约束容量最优的。

表 2 总结了常见广义 AMP 类接收机, 其中 T 为接收机最大迭代次数。值得注意的是, 由于 Turbo-LMMSE 接收机采用外信息更新方式, 无法实现解量化的非线性操作, 因此无法应用到 GLM。

3.3 数值实验

本节的仿真为基带的链路仿真, 用户数 $M = 500$, 服从



▲图9 基于内迭代的广义AMP类接收机

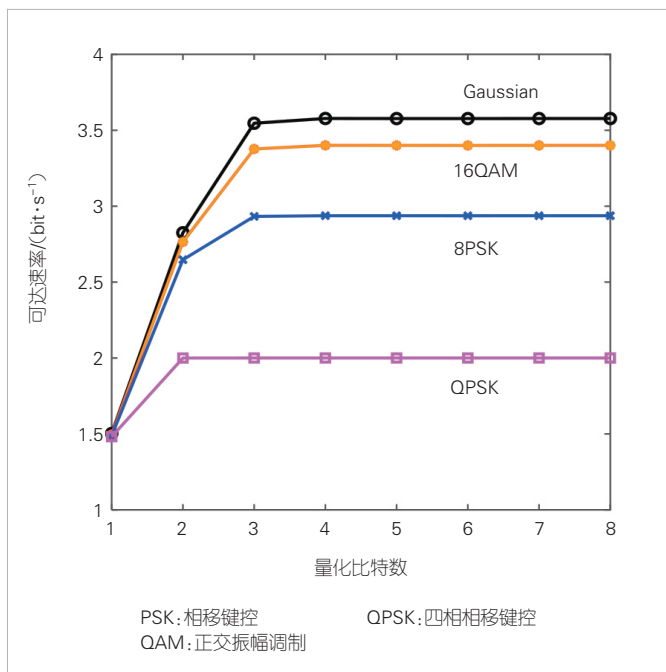
▼表2 MU-MIMO系统下广义AMP类接收机对比

方案/算法	信号	信道矩阵	信息理论极限	计算复杂度
广义AMP ^[14]	任意	独立同分布	最大可达速率	低复杂度 $O(MNT)$
广义记忆AMP ^[16]		酉不变		高复杂度 $O(M^2NT+MNT)$
广义正交AMP ^[15]				

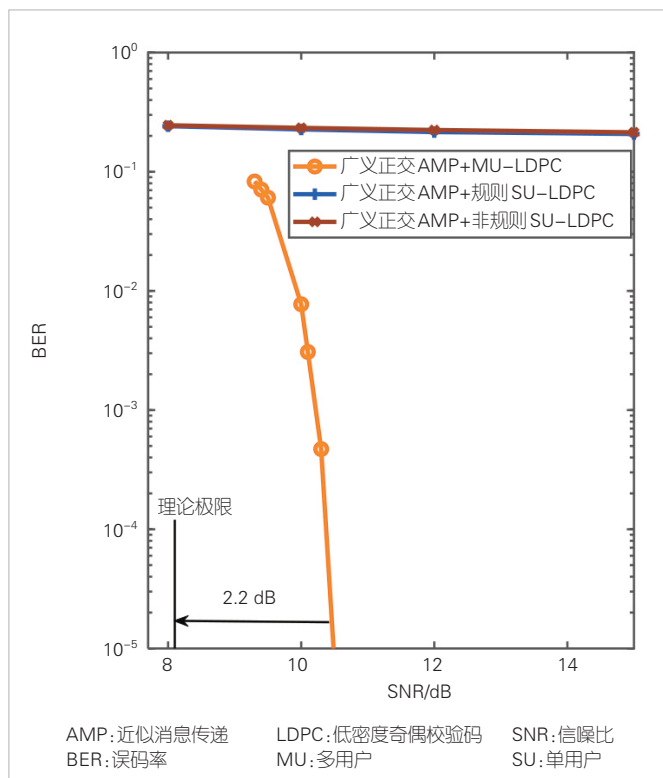
MU-MIMO:多用户多输入多输出 AMP:近似消息传递

均匀分布,发送功率为1。发送端采用MU-LDPC,接收端采用广义正交AMP接收机进行信号检测,信道为一般酉不变信道,具体模型为:信道矩阵 A 的奇异值分解 $A = U\Lambda V^H$,将 A 的特征值 $\{d_i\}$ 设置为: $d_i/d_{i+1} = \kappa^{1/T}, i = 1, 2, \dots, T, \sum_{i=1}^T d_i^2 = N$,其中 $T = \min\{M, N\}$, $\kappa > 1$ 表示 A 的条件数, U 和 V 由独立同分布高斯矩阵的QR分解生成。信道矩阵 A 仅接收端已知,发送端未知。 $Q(\cdot)$ 为逐符号的量化函数,对输入符号的实部和虚部分别进行量化。

当信道矩阵 A 的条件数 $\kappa = 50$ 、接收天线数 $M = 500$,信噪比固定为20 dB时,图10比较了广义正交AMP在不同量化比特下最大可达速率。对于不同的输入信号,采用3 bit量化可以保证广义正交AMP接收机能够达到最大可达速率。图11进一步给出当采用1 bit量化和QPSK调制时,码率为0.4的MU-LDPC与码率为0.4的SU-LDPC(规则和非规则)的BER性能比较。可以看出,在1 bit量化场景下现有的规则和非规则的SU-LDPC已经无法准确恢复信息,而MU-LDPC仍能够准确恢复信息,且距离信息理论极限仅差2.2 dB。



▲图10 广义正交近似消息传递(AMP)不同量化比特的可达速率比较



▲图11 广义正交AMP接收机不同方案BER比较

4 结束语

本文研究了实际通信系统约束下的大规模离散MU-MIMO系统,并提出了低复杂度、信息理论最优接收机和多用户码设计准则。本文不仅证实了AMP、正交AMP和记忆AMP接收机在理想ADC器件下的容量最优性,而且给出了最优多用户码设计准则。并进一步考虑了低分辨率ADC对接收机性能的影响,给出了广义AMP、广义正交AMP和广义记忆AMP接收机的最大可达速率分析和相应的最优码设计准则。此外,数值实验证实其BER性能显著优于SU-LDPC,特别是在恶劣信道和1 bit量化场景下,解决了单用户码无法准确恢复信息的问题。这为应对未来6G复杂场景通信提供了有效的解决方案,有助于提升无线通信系统的效率、可靠性和适应性。

参考文献

- [1] DAI L L, WANG B C, DING Z G, et al. A survey of non-orthogonal multiple access for 5G [C]//Proceedings of IEEE Communications Surveys & Tutorials. IEEE, 2018: 2294–2323. DOI: 10.1109/COMST.2018.2835558
- [2] PING L, LIU L H, WU K Y, et al. Interleave division multiple-access [J]. IEEE transactions on wireless communications, 2006, 5(4): 938–947. DOI: 10.1109/TWC.2006.1618943
- [3] CHI Y H, LIU L, SONG G H, et al. Practical MIMO-NOMA: low complexity and capacity-approaching solution [J]. IEEE transactions on wireless communications, 2018, 17(9): 6251–6264. DOI: 10.1109/TWC.2018.2858222
- [4] YU J L, YUAN Y P, ZHANG B L. 1-bit nonlinear mapping precoder for downlink massive MU-MIMO systems [C]//Proceedings of 30th Wireless and Optical Communications Conference (WOCC). IEEE, 2021: 162–165. DOI: 10.1109/WOCC53213.2021.9603167
- [5] CHERIF M, BOUALLEGUE R. The effect of high power amplifier nonlinearity on MU-Massive mimo system performance over Rayleigh fading channel [C]//Proceedings of 15th International Wireless Communications & Mobile Computing Conference (IWCMC). IEEE, 2019: 1426–1429. DOI: 10.1109/IWCMC.2019.8766449
- [6] KWAK H Y, NO J S, PARK H. Design of irregular SC-LDPC codes with non-uniform degree distributions by linear programming [J]. IEEE transactions on communications, 2019, 67(4): 2632–2646. DOI: 10.1109/TCOMM.2018.2889850
- [7] CHI Y H, LIU L, SONG G H, et al. Constrained capacity optimal generalized multi-user MIMO: a theoretical and practical framework [J]. IEEE transactions on communications, 2022, 70 (12): 8086–8104. DOI: 10.1109/TCOMM.2022.3207813
- [8] GUO D N, SHAMAI S, VERDU S. Mutual information and minimum mean-square error in Gaussian channels [J]. IEEE transactions on information theory, 2005, 51(4): 1261–1282. DOI: 10.1109/TIT.2005.844072
- [9] LIU L, LIANG C L, MA J J, et al. Capacity optimality of AMP in coded systems [J]. IEEE transactions on information theory, 2021, 67(7): 4429–4445. DOI: 10.1109/TIT.2021.3083748
- [10] MA J J, PING L. Orthogonal amp [J]. IEEE access, 2017, 5: 2020–2033. DOI: 10.1109/access.2017.2653119
- [11] LIU L, LIANG S S, PING L. Capacity optimality of OAMP in coded large unitarily invariant systems [C]//Proceedings of IEEE International Symposium on Information Theory (ISIT). IEEE, 2022: 1384–1389. DOI: 10.1109/ISIT50566.2022.9834360
- [12] LIU L, HUANG S Q, KURKOSKI B M. Memory AMP [J]. IEEE transactions on information theory, 2022, 68(12): 8015–8039. DOI: 10.1109/TIT.2022.3186166
- [13] LIU L, CHI Y H, YUEN C, et al. Capacity-achieving MIMO-NOMA: iterative LMMSE detection [J]. IEEE transactions on signal processing, 2019, 67(7): 1758–1773. DOI: 10.1109/TSP.2019.2896242
- [14] RANGAN S. Generalized approximate message passing for estimation with random linear mixing [C]//Proceedings of IEEE

International Symposium on Information Theory Proceedings. IEEE, 2011: 2168–2172. DOI: 10.1109/ISIT.2011.6033942

- [15] LIU L, CHI Y H, LI Y, et al. Achievable rates of generalized linear systems with orthogonal/vector AMP receiver [J]. IEEE transactions on signal processing, 2023, 71: 4116–4133. DOI: 10.1109/TSP.2023.3328064
- [16] TIAN F Y, LIU L, CHEN X M. Generalized memory approximate message passing for generalized linear model [J]. IEEE transactions on signal processing, 2022, 70: 6404–6418. DOI: 10.1109/TSP.2022.3213414

作者简介



陈学辉，西安电子科技大学信息与通信工程学院在读硕士研究生；主要研究方向为近似消息传递类算法、多载波调制技术等。



池育浩，西安电子科技大学华山准聘副教授、IEEE ICCT2023 移动与无线网络分会联合主席；研究方向为多用户编码与迭代接收机设计、多载波调制技术；主持多个国家自然科学基金青年项目、教育部“春晖计划”项目、陕西省自然科学基金基础研究计划青年项目；获得第30届中国电子学会信息论大会（CIEIT 2023）最佳海报奖；发表论文20余篇。



刘雷，浙江大学百人计划研究员、博士生导师，华为启真优秀青年学者，IEEE 和中国通信学会高级会员，信息论旗舰会议 IEEE ITW 2021 出版联合主席，2020 年度《IEEE Transactions on Communications》期刊模范审稿人，是国际上最早展开记忆消息传递研究的学者之一；长期从事消息传递理论与算法，及其在信号处理、无线通信、信息论及编码领域的应用研究；主持国家自然科学基金优秀青年项目（海外）和青年项目、中兴通讯产学研基金项目，担任国家自然科学基金重大项目子课题负责人、海创基金项目课题负责人；获中国电子学会第30届信息论学术年会青年新星奖和最佳海报奖；发表论文50余篇。

面向卫星通信与导航的下一代多址接入



Next Generation Multiple Access for Satellite Communication and Navigation Networks

侯天为/HOU Tianwei, 关达/GUAN Da, 孙昕/SUN Xin

(北京交通大学, 中国 北京 100044)
(Beijing Jiaotong University, Beijing 100044, China)

DOI: 10.12142/ZTETJ.202401010

网络出版地址: <http://kns.cnki.net/kcms/detail/34.1228.TN.20240226.1524.009.html>

网络出版日期: 2024-02-26

收稿日期: 2023-12-18

摘要: 面向卫星通信与导航 (SCAN) 系统多样复杂的应用场景, 下一代多址接入 (NGMA) 技术可以提供更高的通信速率、更多的接入数量和更广的覆盖面积。面向 SCAN 系统, 对基于多功能 NGMA、基于多资源 NGMA 和基于多技术 NGMA 展开研究, 分析了现有的 NGMA 技术对 SCAN 系统的性能增益。最后, 总结 SCAN-NGMA 面临的挑战, 展望了未来发展的方向, 为 SCAN-NGMA 架构的进一步演进提供了参考依据。

关键词: NGMA; SCAN; 多功能 NGMA; 多资源 NGMA; 多技术 NGMA

Abstract: In the context of diverse and intricate application scenarios within satellite communication and navigation (SCAN) systems, next generation multiple access (NGMA) technology offers the potential for increased communication rates, expanded access capacity, and broader coverage. It focuses on investigating the application of multi-function-based NGMA, multi-resource-based NGMA, and multi-technology-based NGMA within the SCAN system. It also analyzes the performance enhancements that existing NGMA technology can bring to the SCAN system. Furthermore, this paper concludes by summarizing the challenges inherent in SCAN-NGMA system. It also looks ahead to the future development direction, providing valuable insights that serve as a reference for the ongoing evolution of SCAN-NGMA architecture.

Keywords: NGMA; SCAN; multi-function NGMA; multi-resource NGMA; multi-technology NGMA

引用格式: 侯天为, 关达, 孙昕. 面向卫星通信与导航的下一代多址接入 [J]. 中兴通讯技术, 2024, 30(1): 55-59. DOI: 10.12142/ZTETJ.202401010

Citation: HOU T W, GUAN D, SUN X. Next generation multiple access for satellite communication and navigation networks [J]. ZTE technology journal, 2024, 30(1): 55-59. DOI: 10.12142/ZTETJ.202401010

多址接入技术作为移动通信系统中的重要组成部分之一, 一直被誉为无线通信“皇冠上的明珠”。随着 5G 的全面商用, 6G 已经开始被研究。为了满足下一代移动通信系统对超高速率、海量接入和多功能复用信号的需求, 下一代多址接入 (NGMA) 技术就显得至关重要^[1]。

随着无线通信系统的发展, NGMA 可以被分为多功能 NGMA (MF-NGMA)、多资源 NGMA (MR-NGMA) 和多技术 NGMA (MT-NGMA)^[2]。在 MR-NGMA 中, 多种技术正在持续挖掘时域、频域、码域、功率域、空域和多普勒域等资源的非正交性。MT-NGMA 主要关注多种先进通信技术与

NGMA 的结合, 例如智能超表面、去蜂窝通信、毫米波通信等^[3-5]。多种先进技术的结合, 可以针对性地提高无线通信系统性能。在 MF-NGMA 中, 空域、功率域等非正交算法使无线信号具有多种不同的功能, 包括通信感知一体化或通信导航一体化等^[6]。

1 面向卫星通信与导航的 MR-NGMA

卫星通信与卫星导航作为 6G 中不可或缺的部分, 具有全天候、全覆盖等优点。与传统地面通信相比, 卫星通信系统具有超高移动性的特点, 在多址接入技术方面具有较大的不同, 受到了广泛的关注。

1.1 正交时频空接入技术

近年来, 随着学界对 6G 技术的研究, 卫星通信和导航

基金项目: 中央高校基础科研业务费重点项目 (2023JBZY012); 国家自然科学基金青年项目 (62201028); 北京市自然科学基金“海淀联合”项目 (L232041); 中国科协青年人才托举项目 (2022QNRC001); 欧盟“玛丽居里学者”项目 (101106428)

技术蓬勃发展。与传统地面移动通信系统相比,卫星移动速度快、多普勒频移高,传统正交频分复用(OFDM)无法实现面向卫星的高可靠通信,亟需更适合卫星通信的方案。

由于卫星通信系统对数据速率、灵活性和可靠性需求不断增加,正交时频空(OTFS)接入技术受到了广泛的关注。相比于OFDM技术,OTFS将时变信道转换为多时域的时延多普勒信道,是一种面向快时变信道场景的解决方案,具有更显著的性能优势^[7]。OTFS的关键技术是将时变多径信道转换为时延多普勒域内的二维信道,再通过时延多普勒域的均衡技术,使一个传输帧的所有符号经历相同的信道增益。相比于在时频域进行均衡,在时延多普勒域均衡具有信道增益矩阵的稀疏性和低变化性等优点。在调制过程中,OTFS系统首先利用辛有限傅里叶逆变换(ISFFT)将收到的时延多普勒域信号转到时频域。然后在时频域中对信号做Heisenberg变换,将时频域信号转变为连续的时域信号。之后对信号做Wigner变换,将连续时域信号转变为离散时域信号。最后对信号进行辛有限傅里叶变换(SFFT),将信号重新变换回时延多普勒域。OTFS技术利用信道的稀疏性,使原始的时延多普勒域信号在时延域和多普勒域上均得到了扩展,相比于OFDM,减小了信道估计的开销^[8]。

根据不同的应用场景,目前OTFS有不同的研究侧重点。

1) 面向毫米波的OTFS系统:相比于5G系统,毫米波系统相位噪声更高,多普勒扩展更大,利用OTFS调制可以优化系统性能。与OFDM的基波波形高度局域化相反,OTFS将信号与载波的乘积扩展到整个平面,以获得高分集增益。在高多普勒扩散和信道状态信息未知的情况下,OTFS的性能要高于OFDM。但是当存在相位噪声时,接收机处理多普勒扩展的能力将受到限制,导致接收机性能因自干扰而降低^[9]。

2) 基于大规模MIMO的OTFS系统:OTFS在大规模MIMO系统中面临的困难是下行信道的信道估计。由于OTFS-MIMO信道具有时延域的普通稀疏性、多普勒域的块稀疏性和角度域的突发稀疏性,故基于三维结构稀疏的信道,OTFS-mMIMO下行信道的信道估计问题可以被代替为稀疏信号恢复问题。但是,如何降低均衡器复杂度、如何提高下行预编码和信道反馈的效率是进一步研究的难点^[10]。

3) 非正交多址接入赋能的OTFS系统:OTFS-NOMA系统同时使用时延多普勒域和功率域,在单移动用户OTFS-NOMA系统中,一个移动用户和多个静止用户共享频谱。移动用户在使用时延多普勒域的同时,可以将时频域中的资源与静止用户共享,通过设计波束成形,可以最大化用户的通信速率。同时,均衡器可以保证在时延多普勒域中不存在干

扰。目前,设计面向多用户的高可靠OTFS-NOMA系统是进一步研究的方向^[11]。

1.2 随机接入技术

通过随机接入,用户设备和网络之间可以建立无线链路,取得与基站的上行同步,获取上行资源,从而提高系统资源利用率,节约终端功耗。为了满足无线通信系统的发展要求,下一代随机接入技术应增加随机接入的用途,更新导频发送机制,扩大覆盖范围和加快波束故障恢复过程。此外,由于地面终端和卫星之间的传播时延大,传统握手协议的效率低,迫切需要新型随机接入技术来解决这些问题。目前对于随机接入的研究分为以下几种:

1) 潜在活跃系统中的免授权大规模随机接入:在用户具有零星访问特点的卫星通信系统中,每个时间只有一个活跃的用户子集,为每个设备分配唯一的导频会浪费无线通信资源,降低网络效率。采用免授权的随机接入方式,上行用户直接向卫星发送信息可以解决握手协议带来的巨大时延问题。针对导频碰撞的问题,我们可以利用导频域非正交多址接入技术,将不同的非正交导频分配给多个设备从而避免导频碰撞,提高频谱效率,降低中断概率^[12]。

2) 面向卫星网络的全球覆盖随机接入:卫星网络表现出超大链路距离的特征,使得为固定地面网络拓扑设计的随机接入协议失效。为此,使用基于无模型的随机接入方法,可以通过与非平稳网络环境的交互,提高网络吞吐量和降低平均时延,并可以灵活调整吞吐量-冲突之间的权衡问题^[13]。

3) 面向万物互联(IoE)的随机接入:IoE被定义为将人、组织、数据和事物结合一起的网络,是6G的研究热点之一。其中,物联网(IoT)和设备到设备(D2D)是IoE中的重要研究方向。对于卫星IoT,复杂的情况和长距离的传输使得目前的随机接入方案无法适用。在边缘卫星节点上,使用训练序列填充的多载波系统可以避免同步失败,同时可利用训练序列来检测设备活动和估计信道。利用地面-卫星链路固有的稀疏性和IoT终端的零星流量特征,随机接入使用正交近似消息传递多测量向量算法来估计信道时延和终端的活跃情况,再利用旋转不变性等技术来增强信道估计,但面向卫星的随机接入尚未关注动态变化网络^[14]。对于D2D网络,用户设备可以从多天线维度方面入手,利用正交的随机波束低时延方法进行随机接入。通过该方法,用户设备可以在信道状态信息未知的情况下,从不同设备中识别各自的随机接入申请,减少波束成形的时延和计算复杂度,充分利用空间分集,解决具有相同前导码的信息无法同时解调的问题。未来,我们仍需对握手过程进行简化,降低通信时延^[5]。

针对MR-NGMA的特点，我们在表1中对面向卫星通信的MR-NGMA技术及适用场景进行了总结。

2 面向卫星通信与导航的MT-NGMA

为了符合6G系统的发展需求，多种新兴的无线通信技术和NGMA技术的结合将大幅提高无线通信系统性能。本章重点综述智能超表面赋能和空天地一体化辅助的NGMA技术。

2.1 RIS赋能的SCAN-NGMA系统

智能超表面（RIS）技术可以主动修改无线信道环境，成为下一代无线通信研究中的关键技术和研究热点。RIS具有易于部署、绿色无污染、稳定性强和频谱效率高等优点^[16]。针对单用户通信场景，借助RIS的信号叠加能力，可以增大信号的信噪比、通信速率和频谱效率；针对多用户通信场景，可以利用RIS的被动波束成形提高其他用户的通信速率^[3]。由于使用了MIMO技术，商业5G网络接收到的信干噪比（SINR）普遍在3~23 dB^[17]，如果通过部署RIS使接收信噪比由20 dB提升到23 dB，那么根据香农公式可以计算出整体的通信速率只提升了14.91%。而对于SINR普遍在-20 dB的卫星通信场景中，若将接收信噪比由-20 dB提升到-17 dB，那么整体的通信速率将提升99.01%。据此，目前热点研究方向是建立RIS赋能的SCAN-NGMA系统。该系统将高效复用卫星导航和通信信号，利用RIS等关键技术来拓展网络覆盖范围，增大用户接入数量，提升卫星信号接收功率，满足通信系统对信道容量的需求。此外，RIS和NOMA的联合使用起到了相互促进的效果。NOMA可以利用用户的信道状态信息进行叠加编码，从而达到共享频谱的目的。RIS则可以合理地改变多用户的信道状态信息，进一步增加NOMA的频谱效率和能量效率，实现共赢。

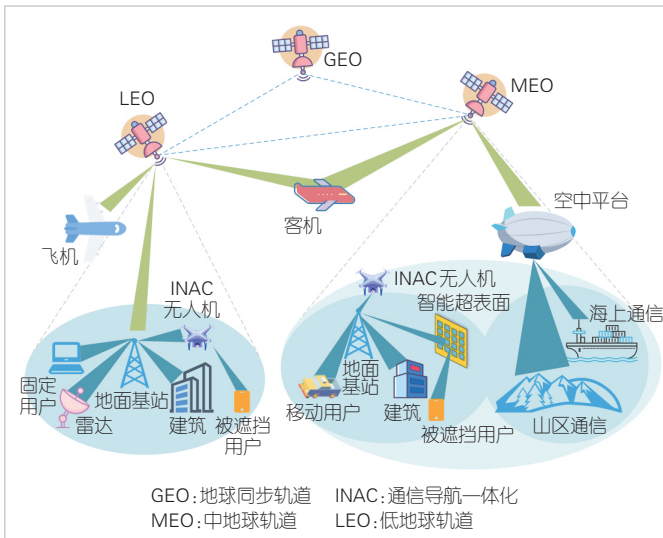
随着材料学技术的发展和RIS技术在应用中遇到的问题，学界提出了同步透射反射RIS（STAR-RIS）。将入射信号转化为透射信号和反射信号进行传输，实现全空间灵活服务，例如实现STAR-RIS辅助的移动边缘计算等^[18-19]。为了减轻无源RIS（PRIS）在大尺度衰落方面的乘性衰落效应对系统性能带来的影响，学界提出了有源RIS（ARIS）的概念。ARIS在原有RIS功能的基础上增加了信号放大的功能，

可以将系统的通信速率提升量从22%增加到130%，进一步提高系统的通信速率和频谱效率^[20]。此外，ARIS也能够进一步增强NGMA系统的性能。通过智能地优化RIS振子的相位，可以根据不同的优化目标选择将小信号增大，增加NGMA系统的策略灵活性，进一步增大系统的总通信速率。

2.2 空天地一体化辅助的SCAN-NGMA系统

空天地一体化网络（SAGIN）是集成了航天、航空和地面的网络，其技术挑战为异构性、时变性、资源有限性和网络的不平衡性。SAGIN因其扩展覆盖范围和节省资源成本的特性而成为了6G的重点研究方向。图1为SAGIN的架构。

根据运行轨道的高度，环绕地球的卫星可分为地球同步轨道（GEO）卫星、中地球轨道（MEO）卫星和低地球轨道（LEO）卫星，大多数卫星网络是由3种卫星组成的异构多层系统，节点类型和网络边缘变化快。由于MEO和LEO与地面的周期性相对移动，卫星网络的边缘将随着时间的推移不断变化，因此传统的多址接入技术很难适用于庞大的异构多层网络^[21]。针对该问题，目前的研究热点是如何在动态的异构多层网络中提高无线资源的利用效率，并降低计算复杂度。一种方法是利用不同网络层到用户之间的信道增益差使用NGMA技术，再使用RIS技术作为辅助，解决接收信号功率过小的问题^[22]。



▲图1 空天地一体化网络(SAGIN)架构示意图

▼表1 面向卫星通信的多资源-下一代多址接入(MR-NGMA)技术及其适用场景

技术	域	资源种类	适用场景
传统多址接入技术	正交频分多址	时域、频域	时间、频率
下一代多址接入技术	非正交多址接入	功率域	时间、频率、功率
	正交时频空接入	时延-多普勒域	时延、多普勒频移
	随机接入	时域、频域、导频域	时间、频率
			具有稀疏性的物联网

无人机是SAGIN中重要的组成部分,与地面基站相比,在移动性、能量限制、用户分布、大尺度衰落信道和小尺度衰落信道等方面有所不同。同时,无人机在通信和导航中可以扮演不同的角色。在通信系统中,无人机既可以被视为移动的基站,为被遮挡的用户或缺少蜂窝的地区提供服务,也可以被视为空中移动用户,接收卫星发送的下行信息。在导航系统中,无人机可以被视为导航服务提供商,为用户的导航服务提供额外的链路,也可以作为导航服务的用户,通过导航信息获取自己的位置并完成移动。无人机的三维移动性和有限能源是影响无人机组网的关键因素,如何实现更高的频谱效率和能源效率是基于无人机的空天地一体化系统的研究重点^[23]。我们可以使用NGMA技术,将不同信道状态的用户根据叠加编码原理进行分组,利用连续干扰消除的技术提高系统总通信速率,增大系统频谱效率^[24]。

3 面向卫星通信与导航的MF-NGMA

在NGMA技术中,实现多功能信号复用是一个广受关注的研究方向。利用多种新兴的无线通信技术,可实现通信、导航、感知等多功能复用,大幅节约频谱资源和能量资源。

3.1 通信导航一体化技术

传统的卫星通信系统和卫星导航系统相互独立,需分别部署独立的硬件系统^[25]。而通信导航一体化(INAC)技术则可以使通信系统和导航系统共享一套硬件,具有总体积小、重量轻、功耗低等优势,理论价值和工程前景较好。目前的研究方向有面向低轨卫星系统和在中轨卫星系统的通信导航一体化技术^[26]。

1) 面向低轨卫星的INAC技术:得益于低轨卫星普遍具有的超宽带通信能力和较高通信速率的特点,INAC不但可以将传统的卫星精密轨道和精密钟差等导航信息传输至用户,还可以同时传输精密时间戳和通信信息至地面用户。此外,低轨卫星与地面用户间可采用双向时间同步,消除时延对时间同步精度的影响,提高时间偏移的估计精度,进一步提高卫星通导一体化系统的性能。此外,根据导航信息的广播性,可以利用新兴的NGMA技术,将导航信息融入到通信信息中,使得无论在强干扰还是弱干扰的情况下,都能够有效提高频谱效率。

2) 面向中轨卫星的INAC技术:利用新兴的叠加编码技术,可实现导航信号和通信信号的复用。接收机利用连续干扰消除技术,地面用户可顺利接收并分别解调通信信号和导航信号,实现通信导航一体化。但传统卫星导航普遍采用中轨卫星和高轨卫星,无法实现高速率的无线通信。针对中轨

卫星的信道特点,在通导一体化用户附近部署RIS阵列,可以有效增强卫星信号,实现高速率通导一体化系统。此外,在城市峡谷或室内定位中,用户可能无法收到足够的卫星信号。利用RIS可以扩展出额外的卫星-RIS-用户链路,实现卫星通信和卫星导航功能^[27]。NGMA赋能的卫星通信导航一体化技术仍处于研究初期,尚未完全挖掘系统的性能增益。

3.2 通信感知一体化技术

随着无线通信行业的飞速发展,频谱成为稀缺资源。为了实现通信和感知的同频复用,学术界和工业界提出了通信感知一体化(ISAC)技术。在ISAC系统中,通信和传感可以在一个硬件平台上同时实现,从而减少对无线资源的需求。SCAN系统具有时延大和移动性强的特点,这会导致瞬时信道状态信息估计困难,故需要设计基于统计信道状态信息的卫星ISAC系统。此外,还可以利用波束成形技术和信道状态信息对通信和感知两方面进行无线资源分配,以提高能量利用效率和频谱利用效率^[28]。

面向SCAN系统的ISAC的主要研究难点是通信和感知对波束成形的不同要求,在适用于感知系统的高频频段面临着巨大的路径损耗。在这种情况下,传感需要时变的多向扫描波束,而通信需要稳定和精确指向的波束,以实现更大的波束成形增益。传统的单波束方法仅提高了通信的性能,而通过使用两个模拟天线阵列,可以实现感知功能和通信功能的集成,在智能汽车和无人机等平台上具有良好的应用前景,具有重要的未来研究前景^[29]。

4 结束语

在6G中,面向卫星通信与导航的NGMA将在实现万物智联、全球覆盖等功能中起到重要作用。本文提出了MR-NGMA、MT-NGMA和MF-NGMA的概念和特征,分析了面向SCAN系统的学术研究,指出了多种SCAN-NGMA的关键技术和使用场景,为未来面向SCAN-NGMA发展提供了参考。

参考文献

- [1] CALVANESE STRINATI E, BARBAROSSA S, GONZALEZ-JIMENEZ J L, et al. 6G: the next frontier: from holographic messaging to artificial intelligence using subterahertz and visible light communication [J]. IEEE vehicular technology magazine, 2019, 14 (3): 42-50. DOI: 10.1109/mvt.2019.2921162
- [2] GIORDANI M, POLESE M, MEZZAVILLA M, et al. Toward 6G networks: use cases and technologies [J]. IEEE communications magazine, 2020, 58(3): 55-61. DOI: 10.1109/mcom.001.1900411
- [3] HOU T W, LIU Y W, SONG Z Y, et al. Reconfigurable intelligent surface aided NOMA networks [J]. IEEE journal on selected areas in communications, 2020, 38(11): 2575-2588. DOI: 10.1109/jsac.2020.3007039
- [4] ZHANG J Y, FAN J Y, ZHANG J, et al. Performance analysis and optimization of NOMA-based cell-free massive MIMO for IoT [J].

- IEEE Internet of Things journal, 2022, 9(12): 9625–9639. DOI: 10.1109/ijot.2021.3130026
- [5] DAI L L, WANG B C, PENG M G, et al. Hybrid precoding-based millimeter-wave massive MIMO-NOMA with simultaneous wireless information and power transfer [J]. IEEE journal on selected areas in communications, 2019, 37(1): 131–141. DOI: 10.1109/JSAC.2018.2872364
- [6] LI A N, BODANESE E, POSLAD S, et al. A trajectory-based gesture recognition in smart homes based on the ultrawideband communication system [J]. IEEE Internet of Things journal, 2022, 9(22): 22861–22873. DOI: 10.1109/ijot.2022.3185084
- [7] SINHA A K, MOHAMMED S K, RAVITEJA P, et al. OTFS based random access preamble transmission for high mobility scenarios [J]. IEEE transactions on vehicular technology, 2020, 69(12): 15078–15094. DOI: 10.1109/tvt.2020.3034130
- [8] HADANI R, RAKIB S, TSATSANIS M, et al. Orthogonal time frequency space modulation [C]//Proceedings of IEEE Wireless Communications and Networking Conference (WCNC). IEEE, 2017: 1–6. DOI: 10.1109/wcnc.2017.7925924
- [9] HADANI R, RAKIB S, MOLISCH A F, et al. Orthogonal time frequency space (OTFS) modulation for millimeter-wave communications systems [C]//Proceedings of IEEE MTT-S International Microwave Symposium (IMS). IEEE, 2017: 681–683. DOI: 10.1109/mwsym.2017.8058662
- [10] SHEN W Q, DAI L L, AN J P, et al. Channel estimation for orthogonal time frequency space (OTFS) massive MIMO [J]. IEEE transactions on signal processing, 2019, 67(16): 4204–4217. DOI: 10.1109/TSP.2019.2919411
- [11] DING Z G. Robust beamforming design for OTFS-NOMA [J]. IEEE open journal of the communications society, 2020, 1: 33–40. DOI: 10.1109/ojcoms.2019.2953574
- [12] LIU Y X, ZHAO M, XIAO L M, et al. Pilot domain NOMA for grant-free massive random access in massive MIMO marine communication system [J]. China communications, 2020, 17(6): 131–144. DOI: 10.23919/jcc.2020.06.011
- [13] LEE J H, SEO H, PARK J, et al. Learning emergent random access protocol for LEO satellite networks [J]. IEEE transactions on wireless communications, 2023, 22(1): 257–269. DOI: 10.1109/twc.2022.3192365
- [14] YING K K, GAO Z, CHEN S, et al. Quasi-synchronous random access for massive MIMO-based LEO satellite constellations [J]. IEEE journal on selected areas in communications, 2023, 41(6): 1702–1722. DOI: 10.1109/JSAC.2023.3273699
- [15] LEE J H, KIM Y, LEE I G. Random beam-based random access for low-latency device-to-device communication systems [J]. IEEE access, 2020, 8: 79887–79895. DOI: 10.1109/ACCESS.2020.2990448
- [16] LIU Y W, LIU X, MU X D, et al. Reconfigurable intelligent surfaces: principles and opportunities [J]. IEEE communications surveys & tutorials, 2021, 23(3): 1546–1577. DOI: 10.1109/comst.2021.3077737
- [17] ANDREWS J G, BAI T Y, KULKARNI M, et al. Modeling and analyzing millimeter wave cellular systems [J]. IEEE transactions on communications, 2016, 1. DOI: 10.1109/tcomm.2016.2618794
- [18] MU X D, LIU Y W, GUO L, et al. Simultaneously transmitting and reflecting (STAR) RIS aided wireless communications [J]. IEEE transactions on wireless communications, 2022, 21(5): 3083–3098. DOI: 10.1109/twc.2021.3118225
- [19] QIN X T, SONG Z Y, HOU T W, et al. Joint resource allocation and configuration design for STAR-RIS-enhanced wireless-powered MEC [J]. IEEE transactions on communications, 2023, 71(4): 2381–2395. DOI: 10.1109/tcomm.2023.3241176
- [20] ZHANG Z J, DAI L L, CHEN X B, et al. Active RIS vs. passive RIS: which will prevail in 6G? [J]. IEEE transactions on communications, 2023, 71(3): 1707–1725. DOI: 10.1109/tcomm.2022.3231893
- [21] ZHU X M, JIANG C X, KUANG L L, et al. Capacity analysis of multi-layer satellite networks [C]//Proceedings of International Wireless Communications and Mobile Computing (IWCMC). IEEE, 2020: 251–256. DOI: 10.1109/iwcmc48107.2020.9148095
- [22] GUAN D, SUN X, WANG J, et al. RIS-NOMA-aided LEO satellite communication networks [C]//Proceedings of 10th International Conference on Information Systems and Computing Technology (ISCTech). IEEE, 2022: 409–413. DOI: 10.1109/isctech58360.2022.00071
- [23] HOU T W, LIU Y W, SONG Z Y, et al. Exploiting NOMA for UAV communications in large-scale cellular networks [J]. IEEE transactions on communications, 2019, 67(10): 6897–6911. DOI: 10.1109/tcomm.2019.2929806
- [24] HOU T W, LIU Y W, SONG Z Y, et al. UAV-to-everything (U2X) networks relying on NOMA: a stochastic geometry model [J]. IEEE transactions on vehicular technology, 2020, 69(7): 7558–7568. DOI: 10.1109/tvt.2020.2994167
- [25] ZHAO Q C, GONG W F, HOU T W, et al. Global navigation satellite system (GNSS): a reconfigurable intelligent surface (RIS)-aided approach [C]//Proceedings of GLOBECOM 2022 – 2022 IEEE Global Communications Conference. IEEE, 2022: 3162–3167. DOI: 10.1109/globecom48099.2022.10001462
- [26] HOU T W, LI A N. Performance analysis of NOMA-RIS aided integrated navigation and communication (INAC) networks [J]. IEEE transactions on vehicular technology, 2023, 72(10): 13255–13268. DOI: 10.1109/tvt.2023.3277581
- [27] ZHAO Q C, GONG W F, HOU T W, et al. Integrated-navigation-and-communication (INAC): a reconfigurable intelligent surface (RIS)-aided approach [C]//Proceedings of IEEE 97th Vehicular Technology Conference (VTC2023-Spring). IEEE, 2023: 1–6. DOI: 10.1109/vtc2023-spring57618.2023.10199514
- [28] YOU L, QIANG X Y, TSINOS C G, et al. Beam squint-aware integrated sensing and communications for hybrid massive MIMO LEO satellite systems [J]. IEEE journal on selected areas in communications, 2022, 40(10): 2994–3009. DOI: 10.1109/jsac.2022.3196114
- [29] ZHANG J A, HUANG X J, GUO Y J, et al. Multibeam for joint communication and radar sensing using steerable analog antenna arrays [J]. IEEE transactions on vehicular technology, 2019, 68(1): 671–685. DOI: 10.1109/tvt.2018.2883796

作者简介



侯天为，北京交通大学副教授；主要研究领域为下一代多址接入、智能超表面、卫星通信与导航系统等；中国科协青年人才托举对象，欧盟“玛丽居里学者”；先后主持和参与基金项目10余项；已发表论文50余篇。



关达，北京交通大学在读博士研究生；主要研究方向为智能超表面、卫星通信导航一体化。



孙昕，北京交通大学教授；主要研究领域为下一代多址接入、专业移动通信等；先后主持和参与基金项目40余项；已发表论文100余篇。

面向 Critical MTC 的无连接传输



Connection-Free Transmission for Critical MTC

李志岗/LI Zhigang^{1,2}, 袁志锋/YUAN Zhifeng^{1,2},
董展谊/DONG Zhanyi^{1,2}, 李文斌/LI Wenbin^{1,2},
梁楚龙/LIANG Chulong^{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 Technology, Shenzhen 518055, China)

DOI: 10.12142/ZTETJ.202401011

网络出版地址: <http://kns.cnki.net/kcms/detail/34.1228.TN.20240227.1009.004.html>

网络出版日期: 2024-02-27

收稿日期: 2024-01-10

摘要: 关键机器类通信 (Critical MTC) 对时延和可靠性都有极高的要求。支持海量 Critical MTC 的终端是一个巨大挑战。提出了一种面向海量 Critical MTC 终端的无连接传输方案。为了降低时延, 提出了基于竞争的无连接单次传输方案; 为了实现高可靠性, 设计了具有极低导频碰撞概率的稀疏正交多导频。进一步利用大规模多输入多输出 (MIMO) 的增益来提高可靠性以及在空域复用大量用户。仿真结果表明, 提出的方案可以支持海量 Critical MTC 终端, 同时满足低时延和高可靠的严格要求。

关键词: 无连接传输; Critical MTC; 低时延高可靠; 导频碰撞; 稀疏正交导频

Abstract: Critical machine-type communication (Critical MTC) has extremely high requirements for latency and reliability. Supporting massive critical MTC terminals simultaneously is a huge challenge. A connection-free transmission scheme for critical MTC is proposed. To reduce latency, a contention-based connection-free one-shot transmission scheme is proposed. To achieve ultra-reliability, multiple sparse orthogonal pilots with extremely low probability of pilot collision are designed. Furthermore, the gain of massive multiple-input multiple-output (MIMO) is used to improve reliability and multiplex a large number of users in the spatial domain. The simulation results show that the proposed scheme can support a large number of critical MTC terminals while meeting the stringent requirements of low latency and ultra-reliability.

Keywords: connection-free transmission; Critical MTC; ultra reliable low latency; pilot collision; sparse orthogonal pilot

引用格式: 李志岗, 袁志锋, 董展谊, 等. 面向 Critical MTC 的无连接传输 [J]. 中兴通讯技术, 2024, 30(1): 60-67. DOI: 10.12142/ZTETJ.202401011

Citation: LI Z G, YUAN Z F, DONG Z Y, et al. Connection-free transmission for critical MTC [J]. ZTE technology journal, 2024, 30(1): 60-67. DOI: 10.12142/ZTETJ.202401011

国际电信联盟 (ITU) 在 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]。该建议书提出了 6G^[2-5] 六大典型应用场景。这些场景不仅包括 5G 三大典型应用场景的扩展: 增强移动宽带 (eMBB) 扩展到沉浸式通信、大规模机器类通信 (mMTC) 扩展到海量通信、超可靠低时延通信 (URLLC) 扩展到极可靠低时延通信 (HURLLC), 还包括新的三大应用场景: 通智融合、通感一体和泛在连接。其中, mMTC/海量通信旨在支持大规模的物联网设备连接, 满足海量低成本、低功耗设备的连接需

求, 极可靠低时延通信则是为了满足低时延和高可靠性通信的需求, 以确保实时和关键性应用的可靠性和稳定性。

机器类通信 (MTC) 作为 5G 重要的应用场景之一^[6-8], 按需求可以分为两类: mMTC 和关键机器类通信 (Critical MTC)。其中, Critical MTC 可以认为是极可靠低时延通信和海量通信的融合^[9-10]。Critical MTC 对可靠性和时延都有严格的要求, 未来对终端数量也有较高的需求。Critical MTC 可以应用在工业物联网 (IIoT)、智慧交通、远程医疗等领域^[11]。据白皮书预测, Critical MTC 将是 6G 的一个重要应用场, 高效支持海量 Critical MTC 终端是一个巨大挑战^[12]。

1 Critical MTC 能力指标与关键技术

Critical MTC 除了对可靠性和时延具有严格的要求外,

基金项目: 国家重点研发计划项目 (2020YFB1807202); 广东省重点领域研发计划 (2019B010157002)

对终端连接数量也有较高的需求。本节首先介绍 Critical MTC 的关键能力指标, 然后介绍满足这些指标所涉及的关键技术以及存在的问题。

1.1 Critical MTC 能力指标

Critical MTC 主要涉及可靠性、时延以及连接密度3个能力指标。在5G新空口(NR)中, URLLC可以支持少量的高可靠和低时延终端, 其指标要求为: 对于32字节的数据包, 用户面时延1 ms, 一次传输的可靠性为 $1-10^{-5}$ (99.999%)^[13]。此外, 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。具体来说, 可靠性指标从5G的 $1-10^{-5}$ 提升至 $1-10^{-5} \sim 1-10^{-7}$, 时延指标从5G的空口时延1 ms降至0.1~1 ms, 6G的连接密度从5G的 10^6 个设备/平方公里增加至 $10^6 \sim 10^8$ 个设备/平方公里。根据上述分析, Critical MTC 场景至少要实现 $1-10^{-5}$ 可靠性要求和1 ms的空口时延。随着技术的发展以及在不同领域的广泛部署, Critical MTC 业务的比重将进一步加大, 因此, 还要具备支持海量Critical MTC终端的能力。

1.2 Critical MTC 关键技术

为了满足Critical MTC对高可靠、低时延以及连接数量的要求, 第3代合作伙伴计划(3GPP) R15支持多项物理层(PHY)和媒体接入控制层(MAC)的关键技术, 并在R16和R17得到了进一步增强。本小节将介绍相关的关键技术和规范, 包括灵活的帧结构、免调度传输、重传以及大规模多输入多输出(MIMO)等。

1) 低时延关键技术

传输和调度时间是影响时延的两个重要因素。为了降低传输时延, 5G支持更灵活的帧结构, 以及更灵活的调度单位^[14]。5G Sub-6 GHz支持15/30/60 kHz子载波间隔。更大的子载波间隔可以降低正交频分复用(OFDM)符号的时间, 进而减少时隙(slot)的时间。此外, 5G还支持由1~13个OFDM符号构成的Mini-slot, 可以进一步降低传输时延。

免调度传输可以显著降低调度与传输时延^[15-16]。免调度传输允许用户设备(UE)直接进行数据传输, 而不用事先发送调度请求和等待动态调度, 因此可以降低时延, 节省信令开销。免调度传输分为预配置免调度与竞争免调度。在预配置免调度传输中, 基站为用户预配置周期性的传输资源, 包括时频资源、导频等, 此时用户之间不会发生资源碰撞, 可以认为是非竞争的。因此, 预配置免调度适合周期性的业务。而在竞争免调度传输中, 用户自主地发起信息传输, 无须请求基站分配上行传输资源。竞争免调度传输适合事件/

业务触发的零星、非周期业务, 且资源利用更加灵活和高效。

竞争免调度传输作为一种极简的传输方案, 也存在一些问题。一方面, 在竞争免调度传输中, 用户和基站之间缺少严格的同步, 导致基站接收的用户数据经历较大的时频偏, 这会进一步增加数据符号的畸变, 使信道估计和解调更加困难^[17]; 另一方面, 由于没有基站的调度和协调, 用户的导频信号是自主确定的, 会产生导频碰撞。当用户较多时, 导频碰撞会更加严重, 这会降低传输可靠性。文献[18-20]提出了多种导频设计来降低导频碰撞, 如非正交导频、多导频等。因此, 在存在不受控的导频碰撞、严重的多用户干扰以及时频偏的情况下, 利用竞争免调度传输支持大规模Critical MTC设备依然存在很大的挑战性。

2) 高可靠关键技术

重传作为一种提高系统可靠性的技术, 也在不断演进。传统的混合自动重传请求(HARQ)方案使用停等协议来发送数据。当设备发送一个数据包后, 需要等待基站反馈确认信息, 若接收正确则返回肯定(ACK), 错误则返回否定(NACK)。当设备收到NACK后, 就重新发送上次传输的数据包。传统的重传方案会增加延迟, 而延迟约束下会限制重传的次数, 从而限制可靠性的提高。在此基础上, 文献[21-23]提出了K次重复(K-repetition)、主动HARQ(Proactive HARQ)等多种重传方案。K次重复可以连续重复传输初始数据包K次, 然后再接收基站的反馈信息。主动HARQ允许用户在连续重传过程中接收基站的反馈信息, 一旦收到ACK, 就可以让重传提前终止, 从而降低传输时延。基站通常需要对重传的数据包进行软合并来提升解调性能, 但这会增加数据处理的复杂度。文献[24]提出了基于串行干扰消除(SIC)的K次重复来支持多个Critical MTC设备。但是, 干扰消除过程会增加接收机的数据处理时延。此外, 重传也会增加设备的功耗。

大规模MIMO也称大规模天线阵列技术, 是5G的核心技术。利用大规模MIMO的分集增益和空间复用增益, 可以显著提高系统谱效和可靠性, 以及在空域复用大量用户^[14,25]。因此, 免调度传输结合Massive MIMO, 非常适合对时延以及可靠性敏感的海量连接场景。

针对上述竞争免调度传输带来的导频碰撞、时频偏问题, 以及复杂的重传过程, 亟待提出一种简单高效的数据传输方法来支持海量的Critical MTC终端。

2 面向Critical MTC的无连接传输方案

基于第1节针对Critical MTC关键能力与关键技术的介

绍,以及存在的问题,我们提出了面向Critical MTC的无连接传输方案。该方案可以满足 $1-10^{-5}$ 可靠性和1 ms时延的严格要求,并且可以同时支持海量的Critical MTC设备。无连接传输可以省去随机接入与资源调度等过程,显著降低时延和信令开销。此外,我们提出了时隙级的一次(One-shot)传输方案,采用更大的子载波间隔。该方案不使用重传,可以进一步降低传输时延,实现1 ms的空口时延。为了满足 $1-10^{-5}$ 高可靠的要求,我们设计了极稀疏独立正交多导频,在无须进行导频信号干扰消除的情况下,可以将导频碰撞概率降至 10^{-5} 以下。同时,利用大规模MIMO的分集增益以及空域复用能力,在提升可靠性的同时,可以复用更多的用户,支持海量Critical MTC终端。极低的导频碰撞概率和大规模MIMO提供的更高可靠性,使得基站只需进行并行的多用户检测,无需干扰消除和迭代检测等复杂耗时的处理流程,可以降低基站数据处理时延。此外,本方案也考虑了无连接传输中由于终端和基站之间的不严格同步带来的时频偏对用户解调的影响。

本节首先介绍提出的基于竞争的无连接传输方案以及系统模型,然后介绍具有极低导频碰撞概率的极稀疏正交多导频设计方法以及相应的导频碰撞概率,最后介绍接收机的数据处理流程。

2.1 无连接传输方案

对于低成本和低功耗的物联网终端来说,为了省电,终端在没有业务时,通常处于空闲态(RRC_IDLE),也可以称为无连接态(Connection-free或Non RRC_CONNECTED)。处于空闲态或无连接态的终端为了传输信息,需要先通过随机接入进入连接态(RRC_CONNECTED)。而随机接入过程需要终端与基站进行多次交互,这无疑会增加传输的功耗、信令开销以及时延。R13规范中引入了去激活态(RRC_INACTIVE)。相比于从空闲态跃迁到连接态,从去激活态跃迁到连接态,可以节省信令开销和时延。此外,处于连接态的终端通常没有专用的上行传输资源,每次传输信息前需要先向基站申请上行传输资源,获得基站的上行资源授权后,才能在基站授权的时频资源上传输信息。可见,对于小包的物联网终端来说,传统上行信息传输变得非常低效。

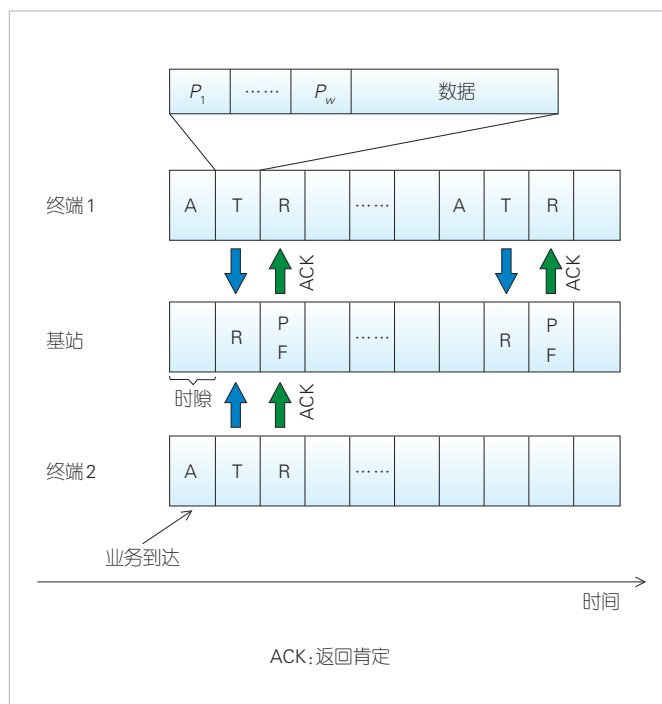
为了降低传输的功耗、信令开销以及时延,我们提出基于竞争的无连接传输:即处于空闲态或无连接状态的终端可以无须事先建立与系统的连接,也无须请求基站分配上行传输资源,而是自主地直接发起信息传输;传输完成后,终端无须释放连接。基站只需配置一个公共的信道,广播通知所

有终端可用的传输资源。基站只需要从这个公共信道的接收信号中,通过多用户检测解码出各个终端的传输信息。可见,这种极简的无连接信息传输,可以极大地降低传输的功耗、信令开销以及时延。

图1为我们提出的基于竞争的无连接传输方案。当业务到达时,终端将自主选择的 W 个导频和数据一起通过最近的可用时隙进行传输,不需要进行重传。基站在下一个时隙会完成多用户检测,并且向用户发送反馈信息。为了确保毫秒级的往返时间,可以采用较大的子载波间隔以减少时隙的持续时间。例如,30 kHz子载波间隔对应的时隙持续时间为0.5 ms,可以确保往返时间在1ms,满足低时延的要求。

2.2 系统模型

考虑大规模MIMO网络,基站配置 M 个天线,服务大量单天线设备。假设 J 个活跃用户同时通过相同的时频资源进行竞争无连接传输。用户从导频序列集合 $\mathbf{P}$ 中自主选择导频序列,与数据一起发送。导频序列集合 $\mathbf{P}$ 由 L 条长度为 L 的正交导频序列构成,即 $\mathbf{P}=[\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_L]$, $\mathbf{p}_l \in C^{1 \times L}$, $l=1, 2, \dots, L$ 。考虑块衰落信道模型,用户和基站间的信道系数可以表示为 $\mathbf{h}_j \in C^{M \times 1}$, $j=1, 2, \dots, J$ 。基站接收到的多用户叠加的导频符号为:



▲图1 基于竞争的无连接传输方案

$$\mathbf{y}_p = \sum_{j=1}^J \mathbf{h}_j \mathbf{p}_j + \mathbf{n}_p, \quad (1)$$

其中, $\mathbf{p}_j$ 、 $\mathbf{y}_p \in \mathbb{C}^{M \times L}$ 、 $\mathbf{n}_p \in \mathbb{C}^{M \times L} \sim \text{CN}(0, \sigma^2)$ 分别表示第 j 个用户选择的导频序列、基站接收到的多用户叠加的导频符号、均值为0方差为 σ^2 的加性高斯白噪声。

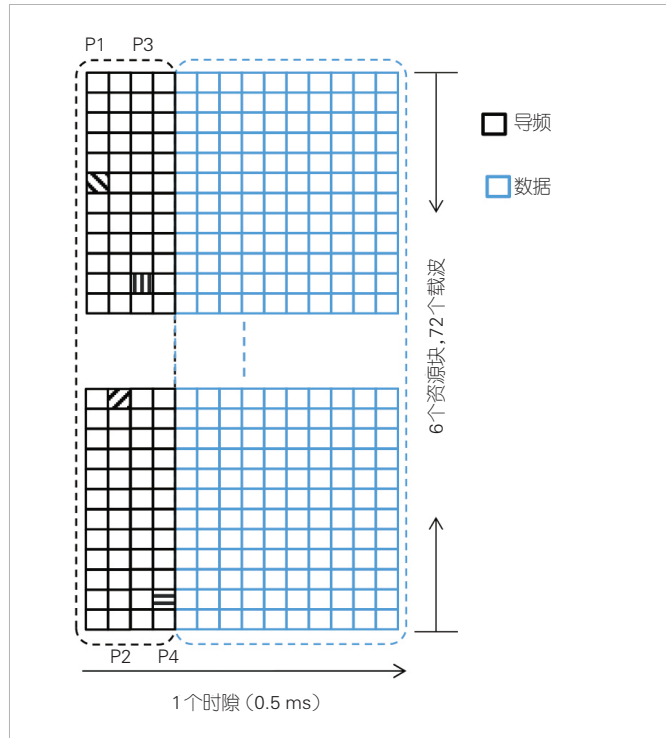
基站接收到的多用户叠加的数据符号为:

$$\mathbf{y}_d = \sum_{j=1}^J \mathbf{h}_j \mathbf{s}_j + \mathbf{n}_d, \quad (2)$$

其中, $\mathbf{s}_j \in \mathbb{C}^{1 \times S}$ 、 $\mathbf{y}_d \in \mathbb{C}^{M \times S}$ 、 $\mathbf{n}_d \in \mathbb{C}^{M \times L} \sim \text{CN}(0, \sigma^2)$ 分别表示第 j 个用户发送的数据符号、基站接收到的多用户叠加的数据符号、均值为0方差为 σ^2 的加性高斯白噪声, S 表示数据符号的长度。

2.3 极稀疏独立多导频设计

在无连接传输方案中, 由于导频是用户自主选择的, 所以不同用户可能会选择相同的导频, 即产生导频碰撞。而高过载即用户很多时, 出现导频碰撞的概率非常高。一旦导频发生碰撞, 基站很难通过导频分离出碰撞用户的信息, 这会显著降低传输的可靠性。为了降低导频碰撞概率, 我们设计了极稀疏独立正交多导频, 如图2所示。传输资源为频域6个资源块 (RB), 时域1个时隙 (包含14个OFDM符号, 每



▲图2 极稀疏独立正交多导频设计

个OFDM符号72个子载波)。以4个导频 ($W=4$) 为例, 4个独立的导频 (每个长度 L 为72) 以时分复用方式映射到时隙的前4个OFDM符号, 即1个导频映射到一个OFDM符号, 导频开销为 $2/7$ 。导频前置的好处是, 基站可以在接收数据的同时, 对导频进行检测, 完成活跃用户识别与信道估计, 降低接收机处理时延。每个72长的导频只有1个非零元素, 该非零元素占一个资源元素 (RE)。这相当于导频集合包含72条稀疏正交导频序列。稀疏正交导频序列集合 $\mathbf{P}$ 可以表示为:

$$\mathbf{P} = \begin{bmatrix} \sqrt{L} & 0 & \cdots & 0 \\ 0 & \sqrt{L} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sqrt{L} \end{bmatrix}_{L \times L}, \quad (3)$$

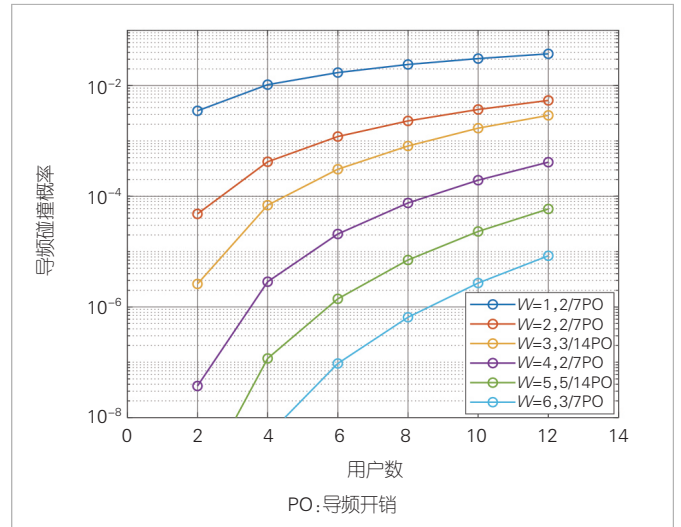
其中 $L=72$, 每一行表示一条稀疏正交导频序列。图2是一种最稀疏的正交导频序列设计, 可以获得数量最多的正交导频序列。当有业务到达时, 用户可以从导频序列集合 $\mathbf{P}$ 中随机选择 W 条序列映射到导频传输资源与数据一起发送。

2.4 导频碰撞概率

极稀疏独立多导频可以显著降低导频碰撞概率。结合大规模MIMO对可靠性的提升, 基站只需要进行一轮多用户检测, 无须复杂耗时的干扰消除和迭代检测过程。在没有干扰消除的情况下, 导频碰撞概率可以表示为^[14]:

$$P_c = (1 - (1 - 1/L)^{I-1})^W. \quad (4)$$

以图2中的时频资源以及导频映射方式为例, 不同导频数量 W 以及不同导频开销 (PO) 下的导频碰撞概率如图3



▲图3 不同导频数量和导频开销下的导频碰撞概率

所示。从图中可以看出，导频碰撞概率达到 10^{-5} 量级时，4个独立导频（2/7导频开销）可以支持8个用户，6个独立导频（3/7导频开销）可以支持12个用户。

2.5 接收机设计

接收机数据处理流程如图4所示，多个独立的导频可以并行处理，这里以其中一个导频的处理流程为例。在无连接传输中，导频可以用于活跃用户检测与信道估计。

由于导频是正交的，可以简单地将接收到的导频符号与导频序列集合 P 中的每一条导频序列进行相关，得到空域信道估计

$$\hat{h}_l = h_l + \frac{n_P P_l^*}{P_l P_l^*}, l = 1, \dots, L. \quad (5)$$

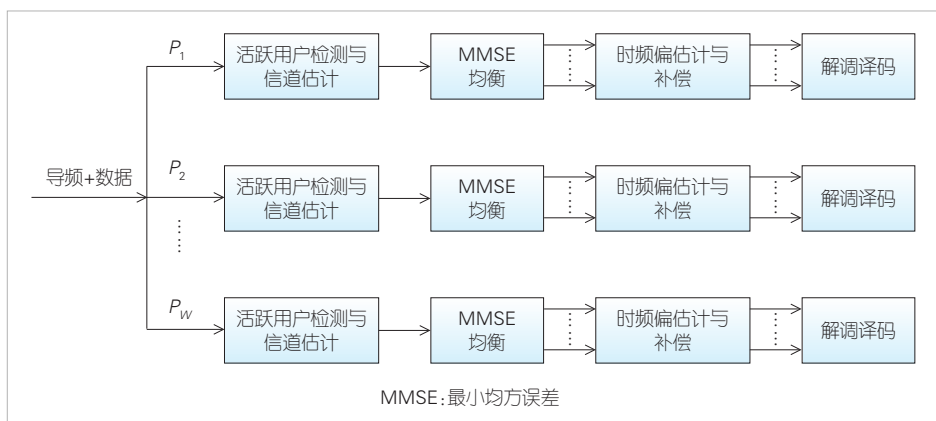
为了识别活跃用户，可以将 $|\hat{h}_l|^2$ 与阈值 T 进行比较。 $|\hat{h}_l|^2 > T$ 即可认为是活跃用户。阈值 T 的大小需要综合考虑漏检和虚警。假设 Q 个活跃用户，信道估计可以表示为 $\hat{H} = [\hat{h}_1, \hat{h}_2, \dots, \hat{h}_Q]$ 。最小均方误差（MMSE）均衡后的数据符号可以表示为：

$$\hat{s}_q = \hat{h}_q^* (\hat{H} \hat{H}^* + \sigma^2 I)^{-1} y_d, q = 1, 2, \dots, Q, \quad (6)$$

其中， $\hat{s}_q$ 表示第 q 个活跃用户的数据符号。对于 Q 个活跃用户，后续处理流程也可以并行进行，即整个接收机处理流程可以认为是两级并行的。这样可以进一步降低接收机处理时延。

为了降低导频碰撞概率，在我们提出的极稀疏正交导频设计中，每个导频只映射到一个RE。导频所在子载波以及附近子载波的信道估计是准确的。对于远离导频的子载波，导频信道估计存在一定偏差。因此公式（6）使用导频信道估计对数据符号进行均衡并不是最优的选择。同时，极简的无连接传输流程也会导致用户和基站之间存在时频偏。也就是说，无线信道以及时频偏会对数据符号 $\hat{s}_q$ 造成畸变，而估计无线信道和时频偏造成的符号畸变超出了稀疏导频的能力范围。

对于小包传输场景，可以采用低阶调制方式，如二进制相移键控（BPSK）、正交相移键控（QPSK）。因此，可以利用低阶调制符号的几何与统计特征来估计无线信道与

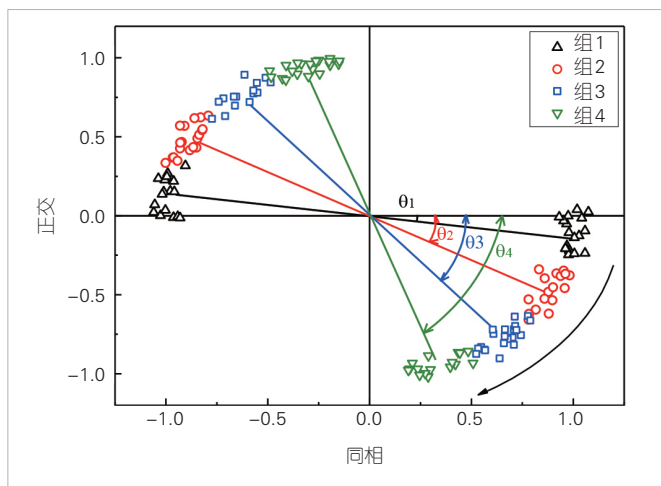


▲图4 接收机数据处理流程

偏。基于此，我们提出了分区匹配（PM）法^[26]。这里以时偏估计为例介绍分区匹配法，考虑正交频分复用系统以及循环前缀（CP）内的时偏。时偏会引起调制数据符号在频域旋转，但是调制数据符号 $\hat{s}_q$ 的几何特征仍然得到保留，如图5所示。为了估计时偏，将传输带宽内的子载波分组，每组中的调制数据符号的平均旋转角度 θ 可通过分区匹配法获得。将这些平均旋转角度进行差分，得到差分角度： $\theta_2 - \theta_1, \theta_3 - \theta_2, \theta_4 - \theta_3, \dots$ 。对差分角度取平均得到 θ_{avg} ，除以分组的子载波数，即可得到两个相邻子载波之间由时偏引起的相位旋转角度。最后，在频域上进行时偏补偿。稀疏导频所在子载波无须进行补偿，并且可以作为其他子载波补偿的参考点。补偿后的数据符号被送到解调器和译码器，通过循环冗余校验（CRC）判断译码是否正确。基站通过译码后的数据即可确定用户身份，并获取业务信息。

3 链路仿真与分析

本节基于MATLAB平台，通过链路仿真验证我们提出



▲图5 分区匹配法估计时偏

的无连接传输方案能够支持海量Critical MTC终端，并满足 $1-10^{-5}$ 可靠性和1 ms时延的指标要求。为确保仿真结果可靠性，我们进行了 3×10^5 次蒙特卡罗仿真。考虑多用户仿真，等效仿真的数据包传输达到 10^6 量级。仿真参数如表1所示。传输资源与图2一致，即频域6个资源块，时域1个时隙。子载波间隔为30 kHz，一个时隙的时间为0.5 ms。导频数量 W 为4~6个，对应导频开销 $2/7 \sim 3/7$ 。基站配置 $M=16$ 或32个阵列天线，用户配置单天线。数据包大小为32字节，载频为5 GHz，信道模型为抽头延时线（TDL-D）模型，时延扩展为30 ns。用户的时偏设置为服从 $[0, CP]$ 独立均匀分布，其中 $CP \approx 2.35 \mu s$ 。用户的频偏设置为服从 $[-250 \text{ Hz}, 250 \text{ Hz}]$ 独立均匀分布。鉴于极稀疏独立多导频可以保证 10^{-5} 量级的导频碰撞概率，加上大规模MIMO可以进一步提升可靠性，因此，我们提出的方案可以只进行单次传输，无须重传，即可满足低时延高可靠的要求。同时接收机只需要进行一轮多用户检测，无需复杂和耗时的干扰消除和迭代检测。

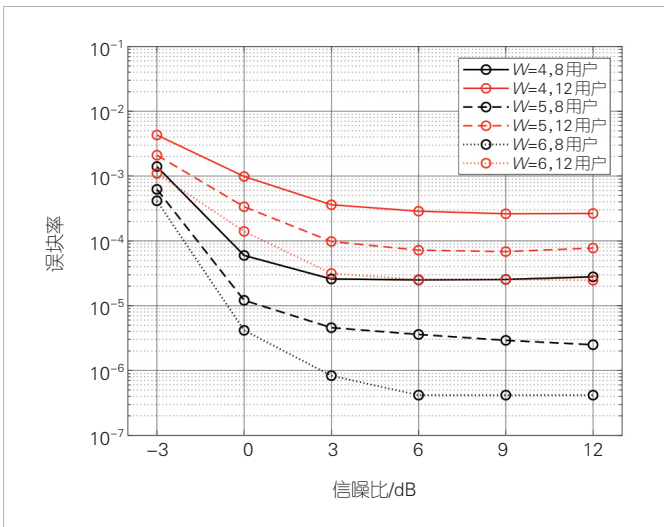
图6为 $M=16$ 时，不同导频开销和导频数量的误块率（BLER）性能。从图中可以看出，在目标BLER= 10^{-5} （可靠性 $1-10^{-5}$ ）时，5个或6个独立稀疏导频都可以支持8个用户。对于6个独立稀疏导频，12用户的BLER并没有降到 10^{-5} ，这与图3中的理想性能还有一定差距。主要原因在于，接收天线数量 M 只有16，规模还不够大。

图7为 $M=32$ 时，不同导频开销和导频数量的BLER性能。6个独立稀疏导频可以支持12个用户，满足BLER= 10^{-5} ，这与图3一致。仿真结果同时表明，我们提出的方案

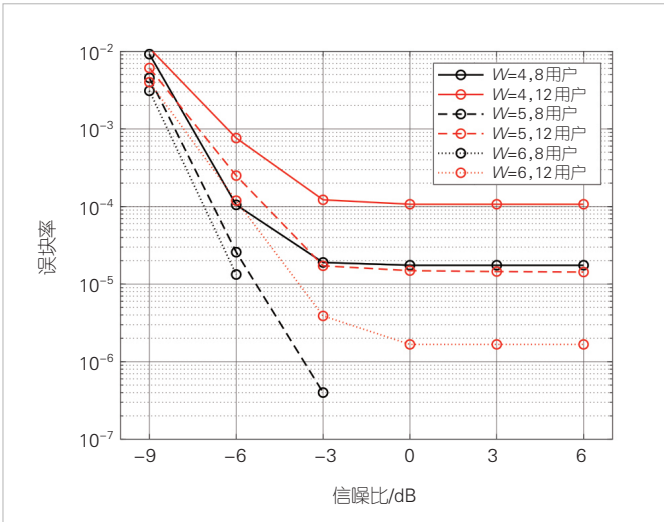
▼表1 仿真参数配置

参数	配置
载频	5 GHz
子载波间隔	30 kHz
传输资源	频域6个资源块,时域1个时隙
用户数	8、12
编码调制方案	BPSK、LDPC
数据包大小	32字节
CRC长度	16比特
导频数量	4、5、6
基站天线配置	16/32 天线阵列
信道模型	TDL-D 30 ns
移动速度	3 km/h
时偏	$[0, CP]$ 均匀分布
频偏	$[-250 \text{ Hz}, 250 \text{ Hz}]$ 均匀分布
接收机算法	MMSE、无干扰消除

BPSK: 二进制相移键控
CP: 循环前缀
CRC: 循环冗余校验
LDPC: 低密度奇偶校验
MMSE: 最小均方误差
TDL-D: 抽头延迟线信道模型



▲图6 $M=16$ 时,不同导频数量和导频开销下的误块率性能



▲图7 $M=32$ 时,不同导频数量和导频开销下的误块率性能

在时频偏场景中依然适用。

图7通过仿真验证了我们提出的方案可以同时支持12个用户满足 $1-10^{-5}$ 可靠性指标的要求。而时延指标主要包括资源调度时延、数据传输时延以及基站数据处理时延。基于竞争的无连接传输，采用更大子载波间隔的时隙级单次传输，可以显著降低资源调度和数据传输时延。基站只需要进行一轮并行的多用户检测，就可以利用图形处理器（GPU）的并行处理能力，降低接收机的数据处理时延。综合以上降低时延的方法可以将空口时延降低至1 ms。

一些Critical MTC应用场景要求海量终端能够进行高可靠传输，但对时延的要求比较宽松（ $>1 \text{ ms}$ ）。基站可以进行干扰消除，通过迭代检测，支持更多用户同时传输。干扰消除不仅针对数据符号，还可以针对导频符号进行，这样可以

进一步降低导频碰撞概率。图8对比了4个独立稀疏正交导频在进行一轮干扰消除和无干扰消除时的导频碰撞概率。从图中可以看出,进行一轮干扰消除,同时接入的用户数超过32时,导频碰撞概率依然可以达到 10^{-5} ;而不进行干扰消除且导频碰撞概率为 10^{-5} 时,4个独立稀疏正交导频仅可以支持6个用户接入。可见导频信号的干扰消除可以显著降低导频碰撞概率,增加同时接入的用户数。

图9对4个独立稀疏正交导频,基站进行一轮干扰消除,不同用户的BLER进行了仿真。其中,基站配置 $M=32$ 天线,信道模型为瑞利衰落信道模型。从图中可以看出,进行一轮干扰消除,可以满足36个用户同时接入,且BLER低于 10^{-5} 。

4 结束语

Critical MTC作为6G的一个重要应用场景,对时延、可

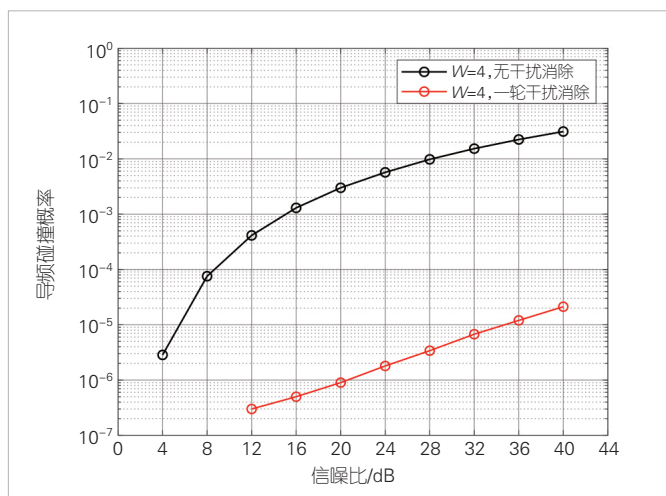
靠性和连接数量有很高的要求。本文中我们介绍了Critical MTC的关键能力指标、所涉及的关键技术以及存在的问题,提出了面向Critical MTC的无连接传输方案,并进行了仿真。仿真结果表明,本文提出的方案可以支持海量Critical MTC终端,同时满足低时延和高可靠的严格要求,在未来通信系统中具有潜在的应用前景。

致谢

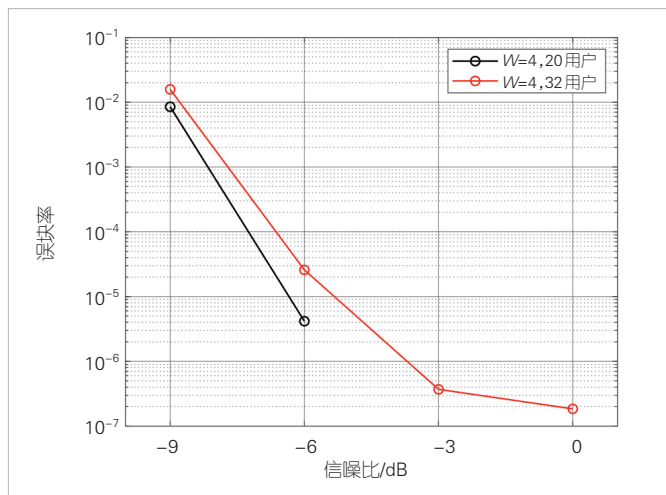
感谢中兴通讯股份有限公司马一华、徐晓景工程师对本文做出的贡献!

参考文献

- [1] ITU. Framework and overall objectives of the future development of IMT for 2030 and beyond [EB/OL]. [2023-06-22]. <https://www.itu.int/md/R19-SG05-C-0131>
- [2] 方敏,段向阳,胡留军. 6G技术挑战、创新与展望[J]. 中兴通讯技术, 2020, 26(3): 61-70. DOI: 10.12142/ZTETJ.202003012
- [3] 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
- [4] 韩书君,董晴,许晓东. 面向6G全场景的多频段协同覆盖扩展技术[J]. 中兴通讯技术, 2022, 28(4): 31-35. DOI: 10.12142/ZTETJ.202204007
- [5] 李福昌,郭希蕊. 6G绿色网络发展趋势和关键技术[J]. 中兴通讯技术, 2023, 29(6): 54-59. DOI:10.12142/ZTETJ.202306009
- [6] 宫诗寻,陶小峰. 5G大规模机器类通信中的传输技术[J]. 中兴通讯技术, 2017, 23(3): 20-23. DOI: 10.3969/j. issn. 1009-6868.2017.03.005
- [7] 陆平,李建华,赵维铎. 5G在垂直行业中的应用[J]. 中兴通讯技术, 2019, 25(1): 67-74. DOI: 10.12142/ZTETJ.201901011
- [8] 陆平,欧阳新志,高雯雯. 5G行业虚拟专网能力提升与实践[J]. 中兴通讯技术, 2022, 28(2): 68-74. DOI: 10.12142/ZTETJ.202202011
- [9] ABDELSADEK M Y, GADALLAH Y, AHMED M H. A critical MTC resource allocation approach for LTE networks with finite blocklength codes [J]. IEEE transactions on vehicular technology, 2020, 69(5): 5598-5609. DOI: 10.1109/TVT.2020.2982391
- [10] POKHREL S R, DING J, PARK J, et al. Towards enabling critical mMTC: a review of URLLC within mMTC [J]. IEEE access, 2020, 8: 131796-131813. DOI: 10.1109/ACCESS.2020.3010271
- [11] MAHMOOD N H, BÖCKER S, MOERMAN I, et al. Machine type communications: key drivers and enablers towards the 6G era [J]. EURASIP journal on wireless communications and networking, 2021, 2021(1): 134. DOI: 10.1186/s13638-021-02010-5
- [12] MAHMOOD N H, BÖCKER S, MUNARI A, et al. White paper on critical and massive machine type communication towards 6G [EB/OL]. [2020-04-29]. <http://arxiv.org/abs/2004.14146.pdf>
- [13] 3GPP. Study on scenarios and requirements for next generation access technologies (release 14): 3GPP TR 38.913 [S]. 2017
- [14] DING J, NEMATI M, POKHREL S R, et al. Enabling grant-free URLLC: an overview of principle and enhancements by massive MIMO [J]. IEEE Internet of Things journal, 2022, 9(1): 384-400. DOI: 10.1109/JIOT.2021.3107242
- [15] JACOBSEN T, ABREU R, BERARDINELLI G, et al. System level analysis of uplink grant-free transmission for URLLC [C]// Proceedings of IEEE Globecom Workshops (GC Wkshps). IEEE, 2017: 1-6. DOI: 10.1109/GLOCOMW.2017.8269137



▲图8 无干扰消除与进行一轮干扰消除的导频碰撞概率对比



▲图9 $M=32$ 时,进行一轮干扰消除时不同用户的误块率(BLER)

- [16] XING S S, XU X Q, CHEN Y, et al. Advanced grant-free transmission for small packets URLLC services [C]//Proceedings of IEEE International Conference on Communications Workshops (ICC Workshops). IEEE, 2019: 1–5. DOI: 10.1109/ICCWork.2019.8757186
- [17] YUAN Z F, LI Z G, LI W M, et al. Contention-based grant-free transmission with extremely sparse orthogonal pilot scheme [C]//Proceedings of IEEE 94th Vehicular Technology Conference (VTC2021-Fall). IEEE, 2021: 1–6. DOI: 10.1109/VTC2021-Fall52928.2021.9625265
- [18] KE M L, GAO Z, WU Y P, et al. Compressive sensing-based adaptive active user detection and channel estimation: massive access meets massive MIMO [J]. IEEE transactions on signal processing, 1809, 68: 764–779. DOI: 10.1109/TSP.2020.2967175
- [19] JIANG H, QU D M, DING J, et al. Multiple preambles for high success rate of grant-free random access with massive MIMO [J]. IEEE transactions on wireless communications, 2019, 18 (10): 4779–4789. DOI: 10.1109/TWC.2019.2929126
- [20] YUAN Z F, LI W M, LI Z G, et al. Contention-based grant-free transmission with independent multi-pilot scheme [C]//Proceedings of IEEE 92nd Vehicular Technology Conference (VTC2020-Fall). IEEE, 2020: 1–7. DOI: 10.1109/VTC2020-Fall49728.2020.9348828
- [21] MAHMOOD N H, ABREU R, BÖHNKE R, et al. Uplink Grant-Free Access Solutions for URLLC services in 5G New Radio [C]//Proceedings of 16th International Symposium on Wireless Communication Systems (ISWCS). IEEE, 2019: 607–612. DOI: 10.1109/ISWCS.2019.8877253
- [22] LIU Y, DENG Y S, ELKASHLAN M, et al. Analyzing grant-free access for URLLC service [J]. IEEE journal on selected areas in communications, 2021, 39(3): 741–755. DOI: 10.1109/JSAC.2020.3018822
- [23] WANG C, CHEN Y, WU Y Q, et al. Performance evaluation of grant-free transmission for uplink URLLC services [C]//Proceedings of IEEE 85th Vehicular Technology Conference (VTC Spring). IEEE, 2017: 1–6. DOI: 10.1109/VTCSpring.2017.8108593
- [24] DING J, CHOI J. SIC aided K-repetition for mission-critical MTC in cell-free massive MIMO [C]//Proceedings of IEEE Conference on Standards for Communications and Networking (CSCN). IEEE, 2021: 167–173. DOI: 10.1109/CSCN53733.2021.9686139
- [25] LI W M, YUAN Z F, MA Y H, et al. Grant-free data-only transmission with correlated antenna array [C]//Proceedings of IEEE Globecom Workshops (GC Wkshps). IEEE, 2021: 1–7. DOI: 10.1109/GCWkshps52748.2021.9681993
- [26] YUAN Z F, HU Y Z, LI W M, et al. Blind multi-user detection for autonomous grant-free high-overloading multiple-access without reference signal [C]//Proceedings of IEEE 87th Vehicular Technology Conference (VTC Spring). IEEE, 2018: 1–7. DOI: 10.1109/VTCSpring.2018.8417836

作者简介



李志岗，中兴通讯股份有限公司技术预研无线高级工程师；主要研究方向为新型多址接入技术、无连接传输。



袁志锋（通信作者），中兴通讯股份有限公司无线算法预研工程师；研究领域涉及信息论、多址、纠错码、自适应算法、MIMO系统和高速软硬件算法。



董展谊，中兴通讯股份有限公司技术预研无线资深专家；主要从事6G预研工作；带领团队完成业务网络架构及相关协议的原型开发工作，并作为首席设计者，开发了验证平台通用驱动程序软件框架和通用eCpri程序软件框架。



李文斌，中兴通讯股份有限公司技术预研无线高级系统工程师；主要从事5G/6G关键技术原型验证相关工作。



梁楚龙，中兴通讯股份有限公司算法部预研工程师；研究方向为信道编码及其在通信系统中的应用；发表论文近40篇。

超大规模天线阵列下的多用户快速波束训练



Fast Beam Training for Multiuser on Extremely Large Aperture Array

司源/SI Yuan^{1,2}, 禹宏康/YU Hongkang^{1,2},
陈艺骞/CHEN Yijian^{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 Technology, Shenzhen 518055, China)

DOI: 10.12142/ZTETJ.202401012

网络出版地址: <http://kns.cnki.net/kcms/detail/34.1228.tn.20240223.1752.002.html>

网络出版日期: 2024-02-26

收稿日期: 2024-01-06

摘要: 天线阵列孔径增加提升了波束赋形增益, 但更窄的波束使得传统的波束训练方法开销成倍增加, 制约了系统支持的用户数目。研究了基于哈希函数的快速、低复杂度新型波束训练方案, 可以用较低的训练开销在支持更多用户的同时进行波束训练, 在此基础上提出了随机及固定两种码本设计方法, 并对波束对齐成功概率进行理论分析。由于推导得到的波束对齐成功概率是一个复杂函数, 无法用于指导码本设计, 为此, 研究提出一个简单的启发式指标用于评估码本生成参数对性能的影响。仿真结果验证了理论分析的正确性, 表明所提波束训练方案能够以低训练开销快速实现多用户波束对齐, 同时具有高准确性。

关键词: 波束训练; 多用户; 哈希函数; 码本设计; 性能分析

Abstract: Large aperture arrays can increase beamforming gain, but lead to narrower beams and multiply the traditional beam training overhead, which restricts the number of users served by the system. A fast and low-complexity beam training scheme based on Hash functions is studied, which can support multi-user simultaneous beam training with low overhead. On this basis, two codebook design methods, random and fixed, are proposed. After that, the beam alignment success probability is derived. Since the derived beam alignment success probability is a complex function and cannot be used to guide codebook design, a simple method is proposed to evaluate the impact of codebook generation parameters on performance. The simulation results verify the correctness of the theoretical analysis and show that the proposed beam training scheme can quickly achieve multi-user beam alignment with low training overhead and high accuracy.

Keywords: beam training; multiuser; Hash function; codebook design; performance analysis

引用格式: 司源, 禹宏康, 陈艺骞. 超大规模天线阵列下的多用户快速波束训练 [J]. 中兴通讯技术, 2024, 30(1): 68-72. DOI: 10.12142/ZTETJ.202401012

Citation: SI Y, YU H K, CHEN Y J. Fast beam training for multiuser on extremely large aperture array [J]. ZTE technology journal, 2024, 30(1): 68-72. DOI: 10.12142/ZTETJ.202401012

采用更高的工作频段和更大的天线阵列是移动通信系统的发展趋势^[1]。当前 5G 系统中已经引入毫米波频段 (FR2 频段)^[2-3], 并将大规模多输入多输出 (MIMO) 作为核心技术 (支持 32 天线端口)^[4]。未来的 6G 系统将进一步扩展至太赫兹通信^[5-9], 采用配置数千甚至上万个天线阵元的极大规模天线阵列 (ELAA) 为用户提供服务。更多的天线数量和更大的阵列孔径能够实现更高的波束赋形增益, 以对抗高频段严重的路径损耗, 保证基站的覆盖距离并大幅提升用户的传输速率。

然而, 波束增益提升也使得波束变窄, 如何以较小的开

销快速实现多用户的接入和波束对齐是一大挑战。传统轮询式波束训练方案依次发送候选波束, 用户可以共用测量导频, 其训练开销与待训练的波束数量成正比。天线阵列规模变大, 波束数量急剧增加, 从而使得轮询式的波束训练方案波束对齐速度变慢。针对这一问题, 分阶段波束训练方案被提出^[10-13], 基站首先发送少量的宽波束, 根据用户反馈缩小波束搜索范围, 最终得到特定宽度的波束。该方案中用户经常不能共用测量导频, 不同用户需要利用时分多址 (TDMA) 的方式进行波束训练, 因此这种方案仅在用户数较少的情况下能够节约开销, 且需要用户多次反馈, 增加了

时延。基于压缩感知的信道估计是另一种能够实现高精度波束对齐的方案^[14-15]，其利用高频信道在角度域的稀疏特性，依靠少量导频开销，通过稀疏信号恢复算法得到估计的信道向量。然而，该方案实现较低开销是以牺牲鲁棒性和提高计算复杂度为代价的，在低信噪比场景下信道估计性能会迅速恶化，并且在多用户场景下，高计算复杂度会使得基站端的处理时延大大增加。

最近，基于随机多波束的快速波束训练方案得到业界广泛关注，其训练波束包含多个传统离散傅里叶变换（DFT）波束，利用哈希函数生成训练波束与DFT波束组合之间的映射关系，因此该方案也称为哈希波束训练^[16-18]。这种波束训练方法本质上是每次发送测量导频时使用多个方向的窄波束进行空间复用，基站从码本中选择一个DFT波束组进行测量导频发送，用户根据测量结果可以推测出最佳波束方向是否在该波束组中。通过多次测量结果的联合判断和交叉校验，用户可以以较低的计算复杂度选出最佳波束方向并反馈给基站。在哈希波束训练中，一个DFT波束组仅对应一份测量导频资源，并且波束组中的波束数量不会影响导频开销，显著缓解了波束数量变多引起的导频开销增大问题；并且，大量用户可以共用同一套测量导频，因此大大节省了导频开销。测量导频的发送波束组对应的码本设计对哈希波束训练方案性能有着重要影响。遗憾的是，现有哈希波束训练方案的码本均采用随机方式生成，未进行专门设计，其性能仍有较大的提升空间。

针对上述问题，本文重点关注哈希波束训练中的码本设计问题，提出随机和固定两种改进的码本设计方案，并对波束对齐成功概率进行理论分析。由于推导出的理论结果形式复杂，本文还提出了具有闭式表达式的启发式指标用于评估哈希码本对性能的影响。根据该指标系统可以快速确定用于生成哈希码本的关键参数，进而提高波束对齐成功率。最后，仿真结果验证了理论分析的正确性，展示出优于已有方案的性能。

1 系统模型

考虑THz点对点多输入单输出（MISO）系统，基站配置 N 天线的均匀线性阵列，服务单天线用户。假设用户位置不变或保持低速移动，信道具有块衰落特性，即在一段时间内可以视为不变。为了表示THz信道的稀疏特性，研究采用几何信道模型^[5]，将收发端信道表示为：

$$\mathbf{h} = \sum_{p=1}^P \alpha_p \mathbf{a}(\theta_p), \quad (1)$$

其中 P 为收发端信道径数， $\mathbf{a}(\theta) = \frac{1}{\sqrt{N}} [1, e^{-j\pi \sin \theta}, \dots, e^{-j\pi(N-1) \sin \theta}]^T$ 表示基站天线阵列的导向矢量， α_p 和 θ_p 分别表示第 p 径的增益和角度。对于THz频段， P 取值通常较小，为2~3^[5]。

研究假设基站进行波束训练的目标是从 N 个相互正交的DFT波束中选择一个最佳波束用于数据传输，令 $\sin \theta_n = -1 + 2n/N$ ，则第 n 个DFT波束的赋形权重可以表示为：

$$\mathbf{f}_n = \mathbf{a}(\theta_n), n = 1, 2, \dots, N. \quad (2)$$

如图1所示，在哈希波束训练方案中，基站依次发送 M 个训练波束，每个训练波束为 L 个DFT波束的叠加，所选 L 个DFT波束的索引通过哈希函数随机生成。本文采用如下方式对其进行建模，定义 $M \times N$ 维的哈希码本 $\mathbf{C}$ ，其中元素 $c_{m,n} \in \{0,1\}$ ， $c_{m,n} = 1$ 表示第 m 个训练波束选中DFT波束 $\mathbf{f}_n$ ，并且有 $\sum_{n=1}^N c_{m,n} = L$ ，则第 m 个训练波束的赋形权重可以表示为：

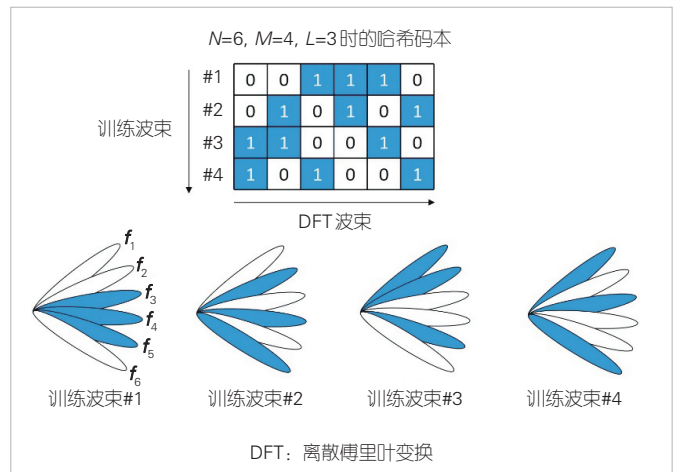
$$\mathbf{w}_m = \frac{1}{\sqrt{L}} \sum_{n=1}^N c_{m,n} \mathbf{f}_n. \quad (3)$$

基于上述模型，用户在第 m 次训练中接收功率可以表示为：

$$y_m = |\mathbf{h}^H \mathbf{w}_m|^2 + n_m, \quad (4)$$

其中 $n_m \sim \mathcal{N}(0, \sigma^2)$ 表示噪声对接收功率的影响。

哈希波束训练的一大优势是用户仅根据测量导频功率 $\{y_1, y_2, \dots, y_M\}$ 即可识别出最佳DFT波束。由于用户和基站预置相同的哈希函数，因此用户根据训练次数索引 m 即可获取哈希码本 $\mathbf{C}$ 。基于此，哈希波束训练采用投票的方式选择最



▲图1 哈希波束训练方案及哈希码本示意图

佳波束,若第 m 次训练中包含第 n 个DFT波束,则将第 m 次训练的接收功率 y_m 累加到DFT波束 n 的测量结果 p_n 中;若DFT波束在某一轮训练中未被选中,则这一轮训练的接收功率不计入该DFT波束的测量结果。第 n 个DFT波束的测量结果可以表示为:

$$p_n = \sum_{m=1}^M c_{m,n} Y_m. \quad (5)$$

最后,用户选择测量结果最大的DFT波束 $f_{\hat{n}}$ 用于数据传输,其中 $\hat{n} = \arg \max_n p_n$ 。

许多已有波束训练方案均可视作哈希波束训练的特例,这些方案的差异仅在于采用的哈希码本。例如,传统轮询式波束训练方案对应 $M = N$ 、 $L = 1$ 、 $\mathbf{C} = \mathbf{I}$;文献[19]的分层波束训练方案对应 $M = 2 \log_2 N$ 、 $L = N/2$ 的哈希码本。因此,哈希码本设计对训练开销以及波束对齐性能有着重要影响。

2 哈希码本设计

本节在文献[17]完全随机哈希码本的基础上,提出部分随机和固定两种改进的码本设计方法。在设计哈希码本时,完全随机方案中训练波束独立的选择DFT波束,对于有限的波束训练次数 M ,各个DFT波束被选中次数 $G_n = \sum_{m=1}^M c_{m,n}$ 差异较大。当用户位于不同位置时,各个DFT波束被成功选中的概率存在较大差异,导致整体波束对齐性能下降。

2.1 所提随机哈希码本

针对完全随机哈希码本存在的问题,所提随机哈希码本各个训练波束通过协作方式选择DFT波束,目标是使各个DFT波束出现次数 G_n 相等。为了实现这个目标,我们限制码本参数 L 能够整除 N ,同时 N/L 能够整除 M 。将训练波束分为 ML/N 组,每组包含 N/L 个训练波束。同一组训练波束通过不放回抽样的方式随机选择DFT波束,每个DFT波束在一个训练波束组内只出现一次,从而有 $G_n = ML/N, \forall n$ 。具体生成哈希码本 $\mathbf{C}$ 的过程如算法1所示。

算法1:所提随机哈希码本设计

输入: DFT波束数 N ,训练波束数 M ,码本生成参数 L

- 1: 初始化 $\mathbf{C} = \mathbf{0}_{M \times N}$
- 2: **for** $m = 1, 2, \dots, M$
- 3: **if** $\text{mod}(m-1, N/L) = 0$, **then** $S = \{1, 2, \dots, N\}$
- 4: 从集合 S 中随机选择 L 个元素构成集合 $T = \{n_1, n_2, \dots, n_L\}$,
- 5: $c_{m,n_i} = 1, 1 \leq i \leq L, S = S - T$

6: **end for**

输出: 所提随机哈希码本 $\mathbf{C}$

2.2 固定哈希码本

对于系统实现而言,采用随机方式生成的哈希码本会导致复杂度增加。例如,基站和用户需要预先配置相同的随机种子用于码本生成。为此,本小节给出一种固定哈希码本设计方法。由于实际应用中码本是预先确定的,因此可以采用离线仿真方式搜索波束训练性能最优的码本。具体来说,研究根据实际场景确定用户分布,得到信道样本;采用3.1小节中的方法生成 X 个所提随机哈希码本 $\{\mathbf{C}_1, \mathbf{C}_2, \dots, \mathbf{C}_X\}$;随后对这 X 个哈希码本在相同信道样本下进行蒙特卡洛仿真,并选择波束对齐性能最优的作为固定哈希码本。与随机生成的哈希码本相比,所提固定哈希码本不仅便于系统实现,还能够更好地匹配实际环境。

3 波束对齐性能理论分析

本节从理论角度分析哈希码本对波束对齐成功概率的影响,为了分析方便,假设信道仅包含视距(LoS)成分,即 $P = 1$ 。对于高频信道来说这一假设是合理的,因为非视距(NLoS)路径增益远远小于LoS径,对接收功率影响较小。此外,不失一般性,我们假设增益 $\alpha = 1$ 。

3.1 波束对齐成功概率的准确表示

将最佳DFT波束下标记作 n^* ,波束对齐成功概率可以表示为:

$$\mathbb{P}\{n^* = \hat{n}\} = \mathbb{P}\left\{\bigcap_{n' \neq n^*} (p_{n^*} > p_{n'})\right\}. \quad (6)$$

根据上式,我们重点分析事件 $p_{n^*} > p_{n'}$ 的概率。对于LoS信道 $\mathbf{h} = \mathbf{a}(\theta)$,仅当 θ 位于DFT波束覆盖范围内时, $|\mathbf{h}^H \mathbf{f}_n|^2$ 具有较大取值。为了简化推导,进一步假设 θ 位于离散格点上,即 $\sin \theta = -1 + 2n/N, n = 1, 2, \dots, N$,此时有 $n^* = n, |\mathbf{h}^H \mathbf{w}_m|^2 = c_{m,n^*}/L$,结合公式(4)和公式(5)有:

$$p_{n^*} = \frac{G_{n^*}}{L} + \sum_{m=1}^M c_{m,n^*} n_m, \quad (7)$$

$$p_{n'} = \frac{K_{n^*,n'}}{L} + \sum_{m=1}^M c_{m,n'} n_m, \forall n' \neq n^*, \quad (8)$$

其中 $K_{n^*,n'} = \sum_{m=1}^M c_{m,n^*} c_{m,n'}$ 表示DFT波束 n^* 和 n' 同时被训练波束选中的次数。根据公式(7)和公式(8),有:

$$p_{n^*} - p_{n'} = \frac{G_{n^*} - K_{n^*,n'}}{L} + \sum_{m=1}^M (c_{m,n^*} - c_{m,n'}) n_m. \quad (9)$$

结合公式 (6), 可以得到事件 $\bigcap_{n' \neq n^*} (p_{n^*} > p_{n'})$, 这意味着方程 $\tilde{\mathbf{A}}_{n^*} \mathbf{n} < \tilde{\mathbf{t}}_{n^*}$ 成立, 其中 $\tilde{\mathbf{A}}_{n^*}$ 和 $\tilde{\mathbf{t}}_{n^*}$ 分别是将 $\mathbf{A}_{n^*} \in \mathbb{R}^{N \times M}$ 和 $\mathbf{t}_{n^*} \in \mathbb{R}^{N \times 1}$ 第 i 列和第 i 行移除得到的结果, 矩阵 $\mathbf{A}_{n^*}$ 的第 (n, m) 个元素为 $c_{m,n} - c_{m,n^*}$, $\mathbf{t}_{n^*} = \left[\frac{G_{n^*} - K_{n^*,1}}{L}, \frac{G_{n^*} - K_{n^*,2}}{L}, \dots, \frac{G_{n^*} - K_{n^*,N}}{L} \right]^T$, 以及 $\mathbf{n} = [n_1, n_2, \dots, n_M]^T$ 。

需要注意的是, 尽管 $\tilde{\mathbf{A}}_{n^*}$ 和 $\tilde{\mathbf{t}}_{n^*}$ 均为已知变量, $\mathbf{n}$ 中元素服从相互独立的高斯分布, 但随机变量 $\tilde{\mathbf{A}}_{n^*} \mathbf{n}$ 的分布在数学上不存在。这是因为哈希波束训练中通常有 $N > M$, 导致协方差矩阵 $\Sigma = \tilde{\mathbf{A}}_{n^*} \tilde{\mathbf{A}}_{n^*}^H$ 并非满秩矩阵。通过加正则化项可以将 $\tilde{\mathbf{A}}_{n^*} \mathbf{n}$ 近似看作服从 $\mathcal{N}(\mathbf{0}, \Sigma + \lambda \mathbf{I})$ 分布, 波束对齐成功概率 $\mathbb{P}\{n^* = \hat{n}\}$ 对应该分布在向量 $\tilde{\mathbf{t}}_{n^*}$ 出的 CDF 函数值。遗憾的是, 计算 $\mathbb{P}\{n^* = \hat{n}\}$ 只能通过数值方式, 无法据此进一步分析码本生成参数对波束对齐性能的影响。

3.2 启发式波束对齐性能指标

为了确定最佳码本生成参数, 即每个训练波束选中 DFT 波束数量 L , 本小节提出一种启发式的评估波束对齐性能的指标。由于固定码本可以通过大量离线仿真优化, 这里重点关注随机码本最佳参数配置。本文选择的指标为随机生成哈希码本时, 最佳 DFT 波束对应测量结果 p_{n^*} 大于任选一个其他波束测量结果 $p_{n' \neq n^*}$ 的概率 $\tilde{p}$ 。根据公式 (9), 在码本确定的情况下, $p_{n^*} - p_{n'} \sim \mathcal{N}\left(\frac{G_{n^*} - K_{n^*,n'}}{L}, (G_{n^*} + G_{n'} - 2K_{n^*,n'})\sigma^2\right)$, 可以得到:

$$\mathbb{P}\{p_{n^*} > p_{n'} | \mathbf{C}\} = \frac{1}{2} \left\{ 1 - \operatorname{erf} \left[\frac{G_{n^*} - K_{n^*,n'}}{L \sqrt{2(G_{n^*} + G_{n'} - 2K_{n^*,n'})\sigma^2}} \right] \right\}, \quad (10)$$

其中, G_{n^*} 以及 $K_{n^*,n'}$ 取值主要受码本参数 L 的影响。接下来我们分别讨论完全随机和所提随机两种码本。

1) 完全随机码本

借助组合数和古典概型理论, DFT 波束 n^* 被选中 g_{n^*} 次的概率为:

$$\mathbb{P}\{G_{n^*} = g_{n^*}\} = C_M^{g_{n^*}} \left(\frac{L}{N} \right)^{g_{n^*}} \left(1 - \frac{L}{N} \right)^{M - g_{n^*}}, g_{n^*} = 0, 1, \dots, M. \quad (11)$$

在 $G_{n^*} = g_{n^*}$ 的条件下, DFT 波束 n^* 和 n' 共同出现 $k_{n^*,n'}$ 的概率为:

$$\mathbb{P}\{K_{n^*,n'} = k_{n^*,n'} | G_{n^*} = g_{n^*}\} = C_{g_{n^*}}^{k_{n^*,n'}} \left(\frac{L-1}{N-1} \right)^{k_{n^*,n'}} \left(1 - \frac{L-1}{N-1} \right)^{g_{n^*} - k_{n^*,n'}}, k_{n^*,n'} = 0, 1, \dots, g_{n^*}. \quad (12)$$

在上述条件下, DFT 波束 n' 被选中 $g_{n'}$ 次的概率为:

$$\mathbb{P}\{G_{n'} = g_{n'} | G_{n^*} = g_{n^*}, K_{n^*,n'} = k_{n^*,n'}\} = C_{M - k_{n^*,n'}}^{g_{n'} - k_{n^*,n'}} \left(\frac{L}{N-1} \right)^{g_{n'} - k_{n^*,n'}} \left(1 - \frac{L}{N-1} \right)^{M - g_{n'}}, g_{n'} = k_{n^*,n'}, k_{n^*,n'} + 1, \dots, M. \quad (13)$$

联合公式 (11) — (13), 可以得到:

$$\mathbb{P}\{G_{n^*} = g_{n^*}, G_{n'} = g_{n'}, K_{n^*,n'} = k_{n^*,n'}\} = \mathbb{P}\{G_{n^*} = g_{n^*}\} \mathbb{P}\{K_{n^*,n'} = k_{n^*,n'} | G_{n^*} = g_{n^*}\} \mathbb{P}\{G_{n'} = g_{n'} | G_{n^*} = g_{n^*}, K_{n^*,n'} = k_{n^*,n'}\}. \quad (14)$$

最终结合公式 (10), 所提性能指标的闭式解为:

$$\tilde{p} = \sum_{g_{n^*}=0}^M \sum_{k_{n^*,n'}=0}^{g_{n^*}} \sum_{g_{n'}=k_{n^*,n'}}^M \mathbb{P}\{p_{n^*} > p_{n'} | \mathbf{C}\} \mathbb{P}\{G_{n^*} = g_{n^*}, G_{n'} = g_{n'}, K_{n^*,n'} = k_{n^*,n'}\}. \quad (15)$$

2) 所提随机码本

对于所提随机码本, $G_n = G = ML/N, \forall n$, 只需分析 $K_{n^*,n'}$ 的分布, 得到:

$$\mathbb{P}\{K_{n^*,n'} = k_{n^*,n'}\} = C_G^{k_{n^*,n'}} \left(1 - \frac{ML - NG}{N} \right) \left(1 - \frac{ML - NG}{N-1} \right) \left(\frac{L-1}{N-1} \right)^{k_{n^*,n'}} \left(1 - \frac{L-1}{N-1} \right)^{G - k_{n^*,n'}}, k_{n^*,n'} = 0, 1, \dots, G. \quad (16)$$

进一步有:

$$\tilde{p} = \sum_{k_{n^*,n'}=0}^G \mathbb{P}\{p_{n^*} > p_{n'} | G_{n^*} = G_{n'} = G\} \mathbb{P}\{K_{n^*,n'} = k_{n^*,n'}\}. \quad (17)$$

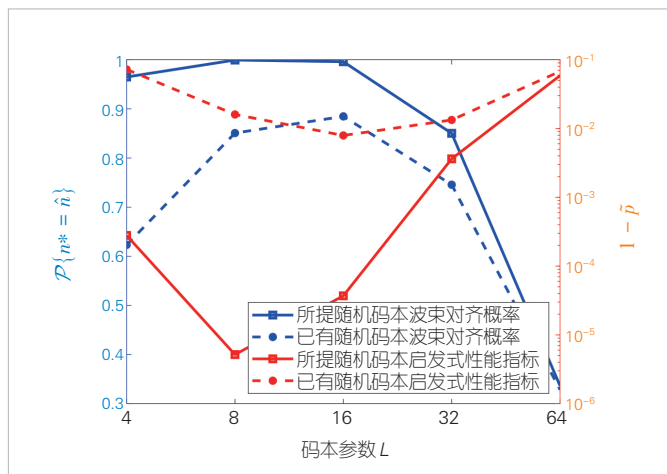
根据上述推导, 公式 (15) 和 (17) 中的启发式性能指标 $\tilde{p}$ 只是 L 的函数, 通过闭式解表达式可以方便计算出最佳码本生成参数 $L^* = \arg \max_L \tilde{p}(L)$ 。

4 仿真分析

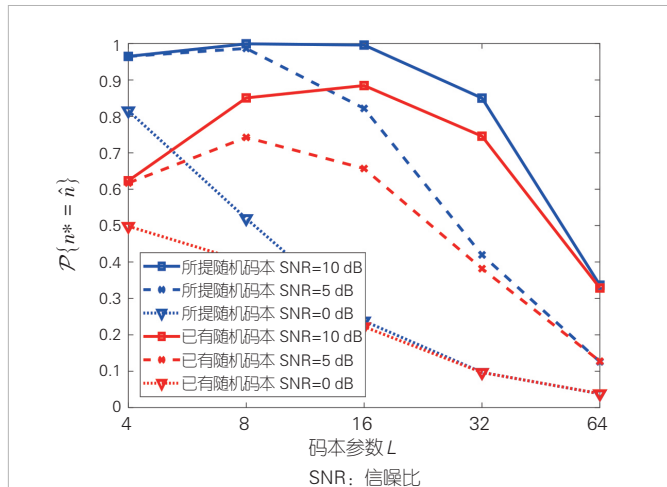
本节对所提哈希波束训练方案进行仿真验证, 并分析了其导频开销。仿真假设基站配备 $N = 128$ 天线, 对应 128 个候选 DFT 波束, 训练波束数 (训练次数) $M = 64$ 。

图2首先对理论分析进行验证,展示了不同随机码本生成参数 L 下,所提随机码本和完全随机码本波束对齐性能。其中,蓝色曲线对应 y 轴左侧刻度,表示波束对齐成功概率 $\mathbb{P}\{n^* = \hat{n}\}$;红色曲线对应 y 轴右侧对数刻度,为了展示清晰,对应启发式指标 $1 - \tilde{p}$ 。仿真假设信噪比 $\text{SNR}=10\text{ dB}$,信道仅包括LoS径,角度位于如第四节所述的离散格点。从图中可以看出,随机码本的生成参数 L 对波束对齐性能有着重要影响,所提启发式指标 $\tilde{p}$ 与波束对齐性能有着近似单调性,即我们可以以最大化 $\tilde{p}$ 为目标获得最佳码本参数 L^* ,使波束对齐准确率最大化。

图3进一步展示了不同信噪比条件下参数 L 对随机码本性能的影响。可以看到,最佳码本参数随信噪比变化。在高信噪比条件下,适中的 L 取值使得波束对齐性能最优。而当信噪比条件变差时, L 的最佳取值逐渐减小。这是因为 L 和信噪比共同决定高斯随机变量 $p_{n^*} - p_n$ 的均值和方差,如公



▲图2 码本参数 L 与波束对齐成功概率和启发式指标的关系

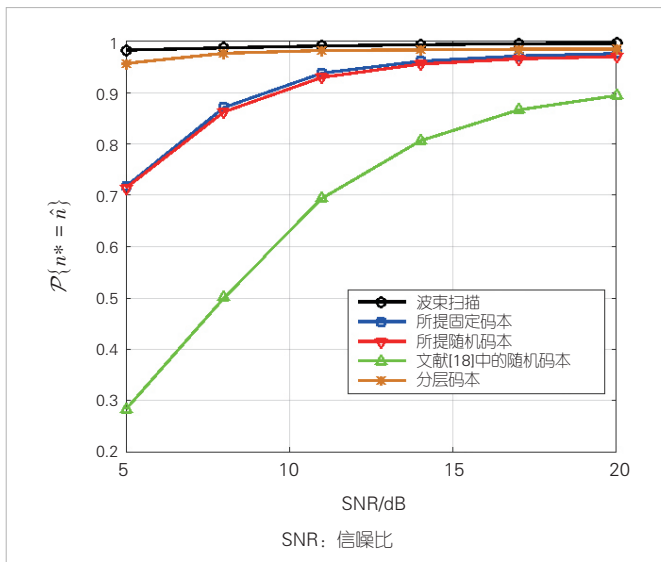


▲图3 不同信噪比下码本参数 L 与波束对齐性能的关系

式(10)所示。此外,当合理设置码本参数时,所提随机码本的波束对齐性能远远优于完全随机码本。

最后,图4展示了实际THz信道场景中不同方案的波束对齐性能。仿真假设一共存在 $P=3$ 条散射路径,包含一条LoS径和两条NLoS径。其中LoS径增益 $\alpha_1 \sim \mathcal{CN}(0,1)$,NLoS径增益 $\alpha_2, \alpha_3 \sim \mathcal{CN}(0,0.01)$,角度 θ_p 在区间 $[-\pi/2, \pi/2]$ 服从均匀分布。仿真采用的固定码本是从1000个随机码本中选出的。从图4中可以看出,在同等训练开销下,所提哈希码本的波束对齐性能远远优于文献[18]中的完全随机码本,在高信噪比下接近波束扫描和每个用户分层波束训练的方案,并且仿真中其训练次数仅有波束扫描的一半。此外,和所提随机哈希码本相比,所提固定哈希码本能够将波束对齐成功率进一步提升约1%,且具有易实现的优势。

表1对比了 P 个用户时随机哈希码本、所提哈希码本、波束扫描、分层波束训练的训练开销。这里的训练开销是指在波束训练的过程中发送导频信号和反馈信号的总数,包括基站发送下行导频的次数(即下行训练次数)和用户进行上行反馈的次数。从表1中可以看出,基于哈希的波束训练开销不随用户数变化,每个用户仅需反馈一次训练结果,适用于多用户接入场景;而分层波束训练的训练开销与用户数 P 成正比,并且每个用户需反馈 $\log_2 N$ 次,用户越多,训



▲图4 不同波束训练方案性能对比

▼表1 波束训练开销对比

方案	下行训练次数	每个用户的上行反馈次数
波束扫描	N	1
所提哈希码本	M	1
随机哈希码本	M	1
分层波束训练	$2P \log_2 N$	$\log_2 N$

训练开销越大, 不适合多用户下的波束训练。

5 结束语

针对超大规模天线阵列波束训练开销过高的问题, 本文中我们研究了基于哈希函数的多用户快速波束训练方案, 重点分析了哈希码本对波束对齐性能的影响。我们提出了随机和固定两种码本设计方法, 并根据理论分析结果得到最佳的码本生成参数。仿真结果表明, 所提方案实现了接近传统轮询方案的性能, 但仅需要一半的训练开销, 并且支持多用户同时训练, 同时所提方案波束对齐准确率远远优于采用完全随机哈希码本的已有方案。

参考文献

- [1] 张跃平. 2022 年度十大天线技术进展 [J]. 中兴通讯技术, 2023, 29 (2): 67–71. DOI: 10.12142/ZTETJ.202302013
- [2] 3GPP. User equipment (UE) radio transmission and reception: 3GPP TS 38.101–1 V18.3.0 [S]. 2023
- [3] 李福昌, 王伟. 5G 高低频组网协同机制与策略 [J]. 中兴通讯技术, 2022, 28(4): 3–6. DOI: 10.12142/ZTETJ.202204002
- [4] 3GPP. Physical layer procedures for data: 3GPP TS 38.214 V18.0.0 [S]. 2023
- [5] DAI L L, TAN J B, CHEN Z, et al. Delay-phase precoding for wideband THz massive MIMO [J]. IEEE transactions on wireless communications, 2022, 21(9): 7271–7286. DOI: 10.1109/TWC.2022.3157315
- [6] 王海明, 陈祎祎. 6G: 跨频段协同通信 [J]. 中兴通讯技术, 2022, 28 (4): 42–43. DOI: 10.12142/ZTETJ.202204009
- [7] 刘利平, 管梦楠, 陈艺骢. 太赫兹技术发展及其应用于 6G 的挑战分析 [J]. 中兴通讯技术, 2021, 27(2): 17–24. DOI: 10.12142/ZTETJ.202102005
- [8] 胡田钰, 李玲香, 陈智. 多频段协同通信的新机遇——太赫兹通信感知一体化 [J]. 中兴通讯技术, 2022, 28(4): 14–18. DOI: 10.12142/ZTETJ.202204004
- [9] 谢峰, 王菲, 刘汉超. 面向 6G 的多频段智能融合组网 [J]. 中兴通讯技术, 2022, 28(4): 25–30. DOI: 10.12142/ZTETJ.202204006
- [10] WANG J Y, LAN Z, PYO C W, et al. Beam codebook based beamforming protocol for multi-Gbps millimeter-wave WPAN systems [J]. IEEE journal on selected areas in communications, 2009, 27(8): 1390–1399. DOI: 10.1109/JSAC.2009.091009
- [11] ZHANG J J, HUANG Y M, SHI Q J, et al. Codebook design for beam alignment in millimeter wave communication systems [J]. IEEE transactions on communications, 2017, 65(11): 4980–4995. DOI: 10.1109/TCOMM.2017.2730878
- [12] CHEN K J, QI C H. Beam training based on dynamic hierarchical codebook for millimeter wave massive MIMO [J]. IEEE communications letters, 2019, 23(1): 132–135. DOI: 10.1109/LCOMM.2018.2881084
- [13] QI C H, CHEN K J, DOBRE O A, et al. Hierarchical codebook-based multiuser beam training for millimeter wave massive MIMO [J]. IEEE transactions on wireless communications, 2020, 19(12): 8142–8152. DOI: 10.1109/TWC.2020.3019523
- [14] RAMASAMY D, VENKATESWARAN S, MADHOW U. Compressive adaptation of large steerable arrays [C]// Proceedings of Information Theory and Applications Workshop. IEEE, 2012: 234–239. DOI: 10.1109/ITA.2012.6181796
- [15] MARZI Z, RAMASAMY D, MADHOW U. Compressive channel estimation and tracking for large arrays in mm-Wave picocells [J]. IEEE journal of selected topics in signal processing, 2016, 10 (3): 514–527. DOI: 10.1109/JSTSP.2016.2520899
- [16] HASSANIEH H, ABARI O, RODRIGUEZ M, et al. Fast millimeter wave beam alignment [C]// Proceedings of the 2018 Conference of the ACM Special Interest Group on Data Communication. ACM, 2018: 432–445. DOI: 10.1145/3230543.3230581
- [17] WANG X, HOU X L, CHEN L, et al. Hash-based fast beam alignment for 6G sub-terahertz MIMO [C]// Proceedings of 27th Asia Pacific Conference on Communications (APCC). IEEE, 2022: 317–322. DOI: 10.1109/APCC55198.2022.9943685
- [18] YOU C S, ZHENG B X, ZHANG R. Fast beam training for IRS-assisted multiuser communications [J]. IEEE wireless communications letters, 2020, 9(11): 1845–1849. DOI: 10.1109/LWC.2020.3005980
- [19] LI Z H, QI C H. Beam training with limited feedback for multiuser mmWave massive MIMO [C]// Proceedings of GLOBECOM 2020 – 2020 IEEE Global Communications Conference. IEEE, 2020: 1–6. DOI: 10.1109/GLOBECOM42002.2020.9348205

作者简介



司源, 中兴通讯股份有限公司技术预研工程师; 主要研究领域为多天线技术、码本设计。



禹宏康, 中兴通讯股份有限公司技术预研工程师; 主要研究领域为多天线技术、近场通信; 发表论文 10 余篇。



陈艺骢, 中兴通讯股份有限公司技术预研资深专家、青年领军人才、深圳市地方级领军人才; 主要研究领域为 6G 技术预研; 发表论文 10 篇, 获授权专利 50 项。

异构大规模分布式网络设计与性能评估



Heterogeneous Large-Scale Distributed Network Design and Performance Evaluation

叶新泉/YE Xinquan^{1,2}, 卢光延/LU Guangyan^{1,2},
陈艺翥/CHEN Yijian^{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 Technology, Shenzhen 518055, China)

DOI: 10.12142/ZTETJ.202401013

网络出版地址: <http://kns.cnki.net/kcms/detail/34.1228.TN.20240306.1159.002.html>

网络出版日期: 2024-02-27

收稿日期: 2024-01-15

摘要: 提出了一种面向6G的异构大规模分布式网络(Het-LDN)。该网络由大量异构的分布式无线节点构成, 具有低成本、低功耗、高性能、易部署和易扩展等特征, 满足未来无线通信的各项需求, 包括实现极致的蜂窝系统容量, 保证超高可行性传输, 支持海量用户同时接入和取得一致性的用户体验。介绍了分层的分布式网络结构与多链路协同传输方案, 并通过仿真实验展示了用户吞吐量的性能优势。认为该网络可成为面向6G的具备强竞争力的商业化解决方案。

关键词: 异构大规模分布式网络; 无线前传; 节点选型; 传输设计; 性能评估

Abstract: A 6G-oriented heterogeneous large-scale distributed network (Het-LDN) is proposed. It consists of a large number of heterogeneous distributed wireless nodes with low-cost, low-power, high-performance, easy-to-deploy, and easy-to-scalable features, which meets the various requirements of future wireless communications, including realizing extreme cellular system capacity, guaranteeing ultra-high feasibility transmission, supporting massive users' simultaneous access, and acquiring consistent user experience. Hierarchical distributed network structure and multi-link cooperative transmission scheme are presented. Simulation experiments demonstrate its performance advantages in terms of user throughput. It is believed that this network can be a highly competitive commercial solution for 6G.

Keywords: heterogeneous large-scale distributed network; wireless fronthaul; node selection; transmission design; performance evaluation

引用格式: 叶新泉, 卢光延, 陈艺翥. 异构大规模分布式网络设计与性能评估 [J]. 中兴通讯技术, 2024, 30(1): 74-81. DOI: 10.12142/ZTETJ.202401013

Citation: YE X Q, LU G Y, CHEN Y J. Heterogeneous large-scale distributed network design and performance evaluation [J]. ZTE technology journal, 2024, 30(1): 74-81. DOI: 10.12142/ZTETJ.202401013

未来的接入网络需要成倍地提升容量, 增加服务用户数目, 并满足针对不同用户的服务质量(QoS), 这对空域自由度提出了更高要求。在多项无线通信技术中, 最具有前景的候选技术为超大规模分布式多输入多输出(MIMO)。该技术通过部署大量的分布式接入节点来构建大规模分布式网络(LDN), 从而实现大量用户的接入和传输。近年来, 该技术已引起业界的重点关注^[1-7]。

LDN是大规模MIMO和分布式MIMO的有机结合, 具备极高宏分集增益和干扰抑制等优点: 1) LDN通过大规模分布式节点的部署, 缩短通信距离, 提供更为均匀的无线信道增益, 带来更为一致且高效的通信体验; 2) LDN通过多

节点的联合信号处理与相干传输, 可以有效地降低多用户(MU)干扰, 提高系统传输速率; 3) LDN提供多条冗余的传输链路, 满足可靠性与灵活性的传输需求, 有助于实现以用户为中心的协作网络设计。因此, 该网络在机场、火车站、体育馆、智慧工厂等场景有着广阔的应用前景^[3]。为了实现通信系统性能与部署成本的良好折中, 异构网(HetNet)被提出^[8]。该网络可由多种不同能力的基站(宏基站、微基站、皮基站、飞基站等)构成, 通过优化的宏微混合组网策略, 在满足特定的网络覆盖需求的前提下, 降低网络的部署成本。此外, HetNet在中心基站整体的协调控制下采用先进的干扰管理方案, 既能保证小区的整体

容量，又能提高小区边缘用户的传输速率，改善用户的整体服务体验。

尽管LDN的性能优势显著，但其不同空间位置的同构站点之间只能通过有线光纤建立连接。有线连接不仅对部署的地理空间环境提出较高要求，而且其大规模同构站点和光纤的部署将会导致不可接受的成本。此外，LDN也会面临着前传与回传容量受限等问题^[4]。虽然HetNet对于上述情况具备一定的弥补特性，但仍需要考虑复杂的多站点干扰管理和资源分配等方案。另外，较高的运营成本和较低的灵活性也是另一制约HetNet大规模应用的重要原因^[9]。因此，本文中，我们将突破LDN和HetNet的限制因素，在接入网侧考虑由两种不同能力的网络节点组成异构大规模分布式网络Het-LDN，并继承上述网络的优势。具体地，第1类网络节点为LDN中的接入点（AP），它与处理单元一同实现宏基站功能；第2类网络节点为具备无线前传能力的各类中继节点（RN），它的大规模部署可扩大网络的覆盖范围，提升容量和数据传输速率。

Het-LDN采用大量低成本的RN代替传统LDN中的部分AP，以实现性能与成本的有效平衡，其具体的特征与优势如下：

- 利用大规模分布式节点联合相干传输，保证热点地区的极致容量需求；
- 提供多条不同的有效传输链路，实现超高可靠性的传输目标；
- 采用动态协作传输方案，避免频繁小区切换，保障一致性体验；
- 使用无线前传替代有线前传，显著降低部署难度；
- 支持灵活的网络参数配置，并根据实际环境与业务需求的变化，按需扩展网络节点；
- 通过异构节点的混合组网，在性能、成本、功耗和复杂度之间取得有效平衡。

为此，可综合考虑两类网络节点选型并联合优化传输算法，构成一套完整的低成本、低功耗、高性能、易部署和易扩展的一体化无线通信解决方案。

1 Het-LDN 系统模型

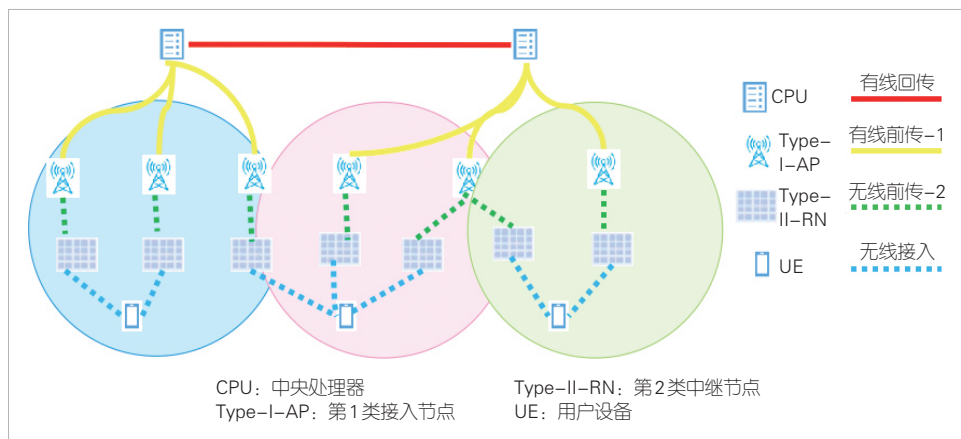
1.1 系统架构与节点选型

Het-LDN 系统架构如图 1 所

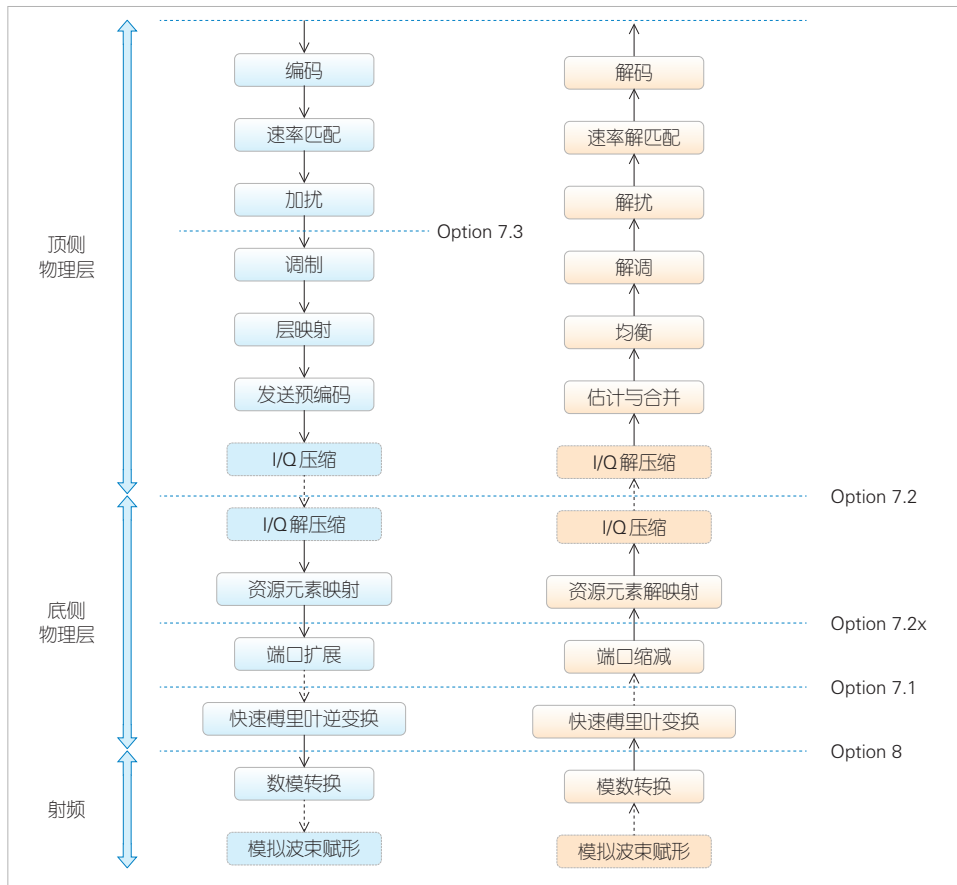
示。Het-LDN 系统由中央处理器（CPU）、Type-I-AP、Type-II-RN 和用户设备（UE）4 类不同的网络节点组成。其中，CPU、Type-I-AP 和 Type-II-RN 至少分别实现 High 物理层（PHY）+ 媒体接入控制（MAC）功能、Low PHY + 射频（RF）功能，以及 RF 功能。不同的 CPU 之间通过有线回传通信（红色实线代表有线回传），CPU 与其所管辖的 Type-I-AP 之间通过有线前传通信（黄色实线代表有线前传-1）。特别地，Type-I-AP 与 Type-II-RN 之间通过无线前传通信（绿色虚线代表无线前传-2）。此外，Type-II-RN 与 UE 还通过无线接入链路通信（蓝色虚线代表无线接入）。

Het-LDN 系统中两类不同网络节点的选型需要综合考虑前传、成本、性能和功耗等多种因素。CPU 与 Type-I-AP 之间的有线前传容量主要由 CPU 最大容量限制、单链路信道质量和并行链路数等因素共同决定。选择合适的 Type-I-AP 在一定程度上可以缓解前传容量受限的问题，且同时在性能、复杂度及成本之间取得有效平衡。Type-II-RN 节点选型则直接影响 Het-LDN 系统的整体性能、成本和功耗。相应地，Type-II-RN 也有多种候选方案。下面我们分别给出 Type-I-AP 与 Type-II-RN 的节点选型过程。

首先考虑 Type-I-AP。根据无线接入网络（RAN）侧所能实现的功能，我们可将现有的接入网功能分离类型分成 Option 8、Option 7.1、Option 7.2x、Option 7.2 和 Option 7.3^[10-11]，不同的接入网功能分离类型如图 2 所示。为了方便起见，上述不同功能分离类型所对应的 Type-I-AP 分别记作 Type-I-1-AP、Type-I-2-AP、Type-I-3-AP、Type-I-4-AP 和 Type-I-5-AP。不同类型的 Type-I-AP 的最显著区别是功能分离的位置，并由此导致前传容量、硬件成本、处理能力、实现复杂度和可扩展性等方面的差异。首先，从前传容量的角度出发，我们可以了解到：越靠近 RF 侧，系统对前



▲图1 Het-LDN系统架构示意图



▲图2 不同接入网功能分离类型示意图

传的要求就越高（如 Type-I-1-AP）；反之，越远离 PHY，系统对前传要求的就越低（如 Type-I-5-AP）。其次，从硬件成本、处理能力和复杂度等方面出发，我们可以发现：Type-I-1-AP 与其他类型的 Type-I-AP 相比，缺少了数字信号处理器（DSP）。因此，Type-I-1-AP 最具备成本优势，但是其只有 RF 功能而不具备快速傅里叶变换（FFT）功能；Type-I-3-AP、Type-I-4-AP 和 Type-I-5-AP 在 AP 侧即可实现本地预编码功能；Type-I-AP 的处理能力与实现复杂度呈正相关关系，如 Type-I-5-AP 的处理能力最强，但其实现复杂度也最高。此外，如果 AP 侧能实现本地预编码，则其具备很好的可扩展性（如 Type-I-3-AP）。根据上述分析，我们可以发现不同类型的 Type-I AP 特征差异显著，其主要特

征对比可参考表 1。进一步地，从表 1 中候选 Type-I-AP 的各项特征，我们可以得出如下关系：

- 前传容量：Type-I-5-AP < Type-I-4-AP < Type-I-3-AP < Type-I-2-AP < Type-I-1-AP；
- 硬件成本：Type-I-2-AP = Type-I-3-AP = Type-I-4-AP = Type-I-5-AP > Type-I-1-AP；
- 处理能力：Type-I-5-AP > Type-I-4-AP > Type-I-3-AP > Type-I-2-AP > Type-I-1-AP；
- 实现复杂度：Type-I-1-AP < Type-I-2-AP < Type-I-3-AP < Type-I-4-AP < Type-I-5-AP；
- 可扩展性：Type-I-3-AP = Type-I-4-AP = Type-I-5-AP > Type-I-1-AP > Type-I-2-AP。

因此，综合考虑上述前传容量、硬件成本、处理能力、实现复杂度和可扩展性等因素，本文中，我们将 Het-LDN 系统中的 Type-I-AP 优选为 Type-I-3-AP

（Option 7.2x），下文简称为 AP。

接下来，我们考虑 Type-II-RN。有线前传不仅对部署环境有一定的要求，还将额外增加部署成本。因此，使用无线前传代替有线前传则可以避免上述的不利因素。最早支持无线前传的 RN 节点包括放大转发（AF）中继和解码转发（DF）中继^[12-13]。其中，前者仅能对接收信号进行直接放大，而后者不仅能对接收信号进行放大，还可以进行编解码处理。网络控制中继器（NCR）^[14]是在不增加过多成本的前提下，以 AF 中继为基础，引入额外网络侧的控制信息，使其具备一定的波束控制能力，从而实现更好的覆盖效果。智能超表面（RIS）^[15]由特定的电磁单元组成，通过施加控制信号来实现对空间电磁波的动态调控，从而主动调整无线环

▼表1 不同候选 Type-I-AP 特征对比

特征	Type-I-1-AP (Option 8)	Type-I-2-AP (Option 7.1)	Type-I-3-AP (Option 7.2x)	Type-I-4-AP (Option 7.2)	Type-I-5-AP (Option 7.3)
功能	射频	射频、快速傅里叶变换	射频、快速傅里叶变换	射频、快速傅里叶变换、第二级预编码、资源元素映射	射频、快速傅里叶变换、第一级和第二级预编码、资源元素映射
容量制约因素	天线端口数	天线端口数	传输层数	传输层数	传输层数
预编码执行侧	CPU	CPU	AP/CPU/AP+CPU	AP/CPU/AP+CPU	AP/CPU/AP+CPU

AP：接入点 CPU：中央处理器

境。进而，根据是否具有功放可将RIS分成被动RIS和主动RIS。为了方便对比不同类型的Type-II-RN特征，我们将AF中继、DF中继、NCR、被动RIS和主动RIS分别记作Type-II-1-RN、Type-II-2-RN、Type-II-3-RN、Type-II-4-RN和Type-II-5-RN。不同的候选Type-II-RN特征对比见表2。进一步地，对比表2中候选Type-II-RN的各项特征，我们可以发现如下关系：

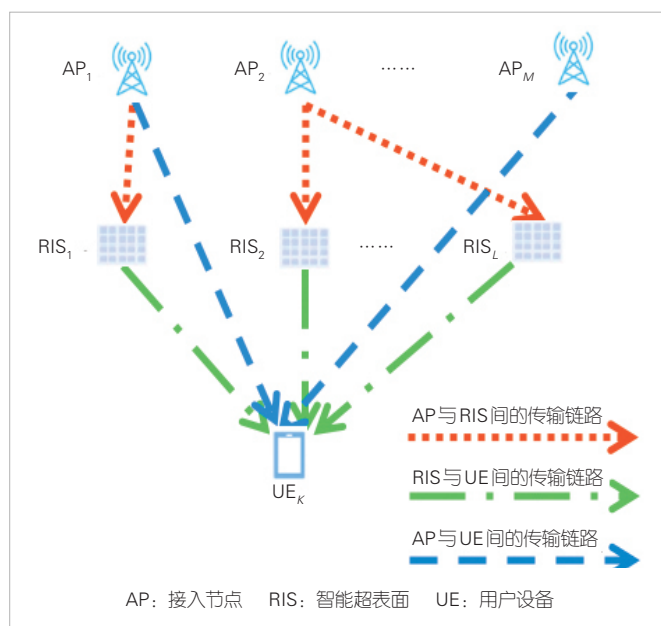
- 硬件成本：Type-II-4-RN<Type-II-5-RN<Type-II-1-RN<Type-II-3-RN<Type-II-2-RN；
- 时延：Type-II-4-RN=Type-II-5-RN<Type-II-1-RN=Type-II-4-RN<Type-II-2-RN；
- 性能：Type-II-2-RN>Type-II-5-RN>Type-II-4-RN>Type-II-3-RN>Type-II-1-RN；
- 功耗：Type-II-4-RN<Type-II-5-RN<Type-II-3-RN<Type-II-1-RN<Type-II-2-RN；
- 实现复杂度，Type-II-4-RN=Type-II-5-RN<Type-II-1-RN<Type-II-3-RN<Type-II-2-RN。

因此，综合考虑上述硬件成本、时延、性能、功耗和实现复杂度等因素，我们将Het-LDN系统中的Type-II-RN优选为Type-II-4-RN（被动RIS）或Type-II-5-RN（主动RIS），下文简称为RIS。

1.2 传输模型

在Het-LDN系统中，每个UE可由一个或多个AP或RIS为其服务，同一时刻一个RIS只能服务一个UE，而一个AP可以同时服务一个或多个RIS。图3给出一种多AP-多RIS的协作传输模型，其中， M 个AP和 L 个RIS分别通过 N_{cd} 个直射路径和 N_{cf} 个反射路径联合为 UE_k 服务。 UE_k 的接收信号模型和干扰信号模型分别如公式（1）和公式（2）所示。其中， x_k 和 x_u 分别是 AP_m 为 UE_k 和 AP_l 为 UE_u 服务时的发送信号， y_k 和 z 为 UE_k 的全部接收信号和全部干扰加噪声信号， z 由 I_1 和 I_2 两部分组成， I_1 为恒定外部干扰（含底噪）， I_2 为

除 UE_k 外其他用户产生的干扰，其数学模型如公式（3）所示， $H_{RU,i}$ 、 $\Phi_{RU,i}$ 和 $p_{RU,i}$ 分别是 UE_k 的第 i 条反射径上RIS与UE之间的信道参数、RIS的调相参数和RIS的调幅参数， $H_{AR,i}$ 、 $W_{AR,i}$ 和 $p_{AR,i}$ 分别是 UE_k 的第 i 条反射径上AP与RIS之间的信道参数、AP的预编码参数和AP的功放参数， H_j 、 W_j 和 p_j 分别是 UE_k 的第 j 条直射径上AP与UE之间的信道参数、AP的预编码参数和AP的功放参数， $H_{RU,a,k}$ 、 $\Phi_{RU,a}$ 和 $p_{RU,a}$ 分别是 UE_u 的第 a 条反射径上RIS $_{aj}$ 与 UE_k 之间的信道参数、RIS $_{aj}$ 为 UE_u 服务时的调相参数和调幅参数， $H_{AR,a}$ 、 $W_{AR,a}$ 和 $p_{AR,a}$ 分别是 UE_u 的第 a 条反射径上服务的AP $_{aj}$ 与RIS之间的信道参数、AP $_{aj}$ 的预编码参数和功放参数， $H_{AR,a,k}$ 是 UE_u 的第 a 条反射径上的服务AP $_{aj}$ 与 UE_k 之间的信道参数， $H_{b,k}$ 、 W_b 和 p_b 分别是 UE_u 的第 b 条直射径上的服务AP $_{bi}$ 与 UE_k 之间的信道参数、AP $_{bi}$ 为 UE_u 服务的预编码参数和功放参数， $N_{cf,u}$ 和 $N_{cd,u}$ 分别是 UE_u 的反射路径数和直射路径数， K 为



▲图3 Het-LDN传输模型示意图

▼表2 不同候选Type-II-RN特征对比

特征	Type-II-1-RN (AF Relay)	Type-II-2-RN (DF Relay)	Type-II-3-RN (NCR)	Type-II-4-RN (被动RIS)	Type-II-5-RN (主动RIS)
功能	只信号放大	可信号放大、可移相、可编解码	可信号放大、受限移相、无编解码	无信号放大、可移相、无编解码	可信号放大、可移相、无编解码
延迟	低	高	低	无	无
功耗	中	较高	中	最低	低
成本	中	较高	中	最低	低

AF: 放大转发
DF: 解码转发

NCR: 网络控制中继器
RIS: 智能超表面

RN: 中继节点

覆盖区域内的所有的 UE 总数。

$$y_k = H_{RU,i} \Phi_{RU,i} P_{RU,i} H_{AR,i} W_{AR,i} P_{AR,i} x_k + \sum_{j=1}^{N_{cd}} H_j W_j P_j x_k + z, \quad (1)$$

$$z = I_1 + I_2, \quad (2)$$

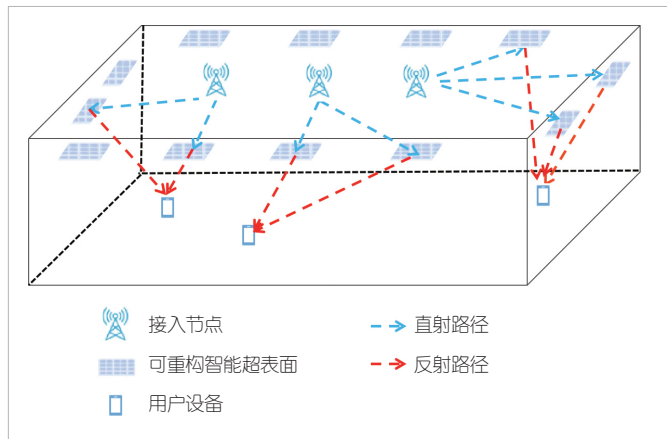
$$I_2 = \sum_{u \neq k}^K \left(\sum_{a=1}^{N_{cf,u}} \left(H_{RU,a,k} \Phi_{RU,a} P_{RU,a} H_{AR,a} W_{AR,a} P_{AR,a} x_u + \right) + \sum_{b=1}^{N_{cd,u}} H_{b,k} W_b P_b x_u \right) \quad (3)$$

2 实验及结果分析

本节中我们基于所搭建的系统级仿真平台，对上述 Het-LDN 系统进行性能评估，其主要包括 Het-LDN 与 LDN 的性能比较和不同参数配置下 Het-LDN 的性能评估。下面我们将给出仿真场景与配置。

2.1 场景与配置

本文中，我们确定 Het-LDN 系统中网络节点的部署策略是评估 Het-LDN 系统级仿真性能的前提。为了降低仿真复杂度，提高系统运行效率，本文的 Het-LDN 系统中暂不考虑 CPU，仅考虑 AP、被动 RIS 和 UE 3 类网络节点的部署，并假定 AP 与 RIS 部署时均只有一层网络结构，且同一层的不同网络节点完全同构。Het-LDN 部署示意如图 4 所示。其中，所有网络节点均限定在一个长方体内（如室内办公室、智慧工厂等），AP 部署在天花板上且沿着天花板中心等间隔分布，RIS 沿着四周墙壁等间隔分布，而 UE 在 2D 空间随机分布。同一类网络节点高度均相同，且有 AP 高度 > RIS 高度 > UE 高度。由于 Het-LDN 的主要应用场景为室内环境，因此 Het-LDN 的主要仿真参数将参考文献[16]中的参数配置，如天线模型、路损模型、视距传播路径概率、穿透损耗和快



▲图 4 Het-LDN 部署示意图

衰模型。另外，考虑到业界尚无广泛统一的 RIS 信道模型，在此我们仅将 RIS 替代为均匀平面阵列（UPA）并进行信道建模。Het-LDN 的详细仿真参数见表 3。下面我们将给出具体性能评估并进行分析。

▼表 3 Het-LDN 仿真参数表

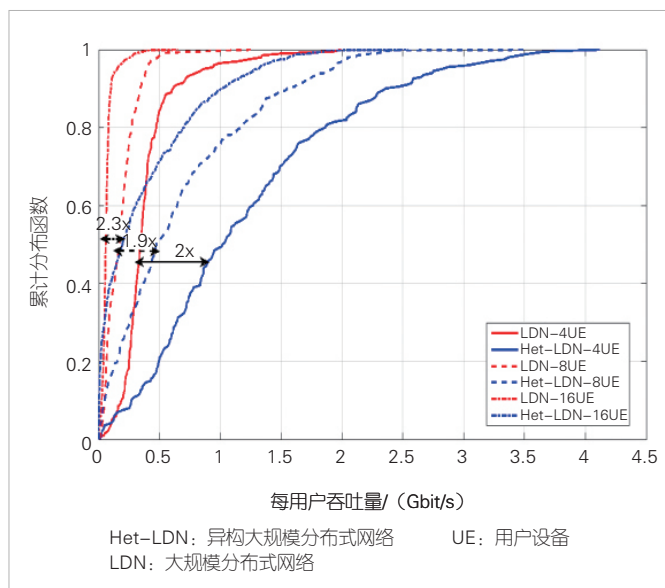
参数	数值
AP 数目	4
RIS 数目	12、24、36、48
UE 数目	4、6、8、10、16
AP 天线数	32
RIS 天线数	512
UE 天线数	4
AP 最大发射功率	23 dBm
AP 高度	6 m
RIS 高度	3 m
UE 高度	1.5 m
天线阵子间距	$\lambda/2$
载波频率	2 GHz
带宽	100 MHz
噪声加外部干扰功率	-120 dBm
区域大小	200 m × 50 m × 6 m
信道模型	3GPP InH

AP: 接入节点 RIS: 智能超表面 UE: 用户设备

2.2 Het-LDN 与 LDN 的性能比较

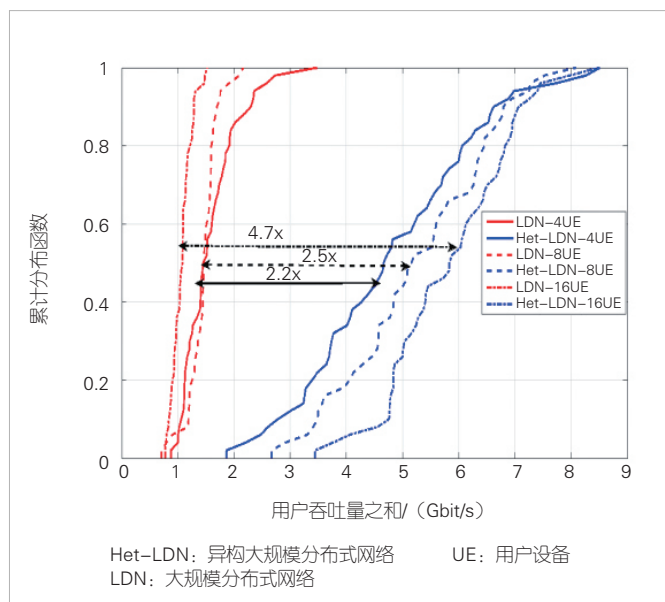
在评估不同参数配置下 Het-LDN 的性能之前，我们首先需要对比 Het-LDN 与 LDN 的性能差异，从仿真实验的角度验证 Het-LDN 相对于 LDN 是否具备优越性。本小节中，Het-LDN 与 LDN 的主要差异体现在网络节点的部署上，即 Het-LDN 相对于 LDN 增加了一层 RIS 节点集合，相对地，LDN 将缺省 RIS 相关参数，而其余的参数在上述两种场景中均相同。具体而言，Het-LDN 场景中的 RIS 数目为 36，UE 数目为 4、8、16，其余仿真参数见表 3。此外，Het-LDN 与 LDN 均采用相同的传输策略，即两种场景均根据长期的信道参数确定协作集且最大协作集规模（协作 RIS 数目或协作 AP 数目）为 2；AP 侧均采用奇异值分解（SVD）预编码^[17]；AP 侧均根据服务 UE 或 RIS 数目等比例分配功率；AP 侧均根据反馈的信道状态信息自适应调整传输层数^[18]；RIS 侧可利用信号传播路径的角度信息和 RIS 面板配置信息确定调相参数^[19]。

图 5 给出了在不同 UE 数目下 Het-LDN 与 LDN 的每用户吞吐量的累计分布函数（CDF）曲线。由于系统增加了 RIS



▲图5 不同UE数目下Het-LDN与LDN的每用户吞吐量的累计分布函数(CDF)曲线

并提供级联的无线通信链路,当UE数目为4时,Het-LDN与LDN相比,中值用户吞吐量提升了2倍;在UE数目为8时,Het-LDN与LDN相比,中值用户吞吐量提升1.9倍;在UE数目为16时,Het-LDN与LDN相比,中值用户吞吐量提升2.3倍。图6为不同UE数目下Het-LDN与LDN的所有用户吞吐量之和的CDF曲线。可以发现,在UE数目为4时,Het-LDN与LDN相比,用户吞吐量之和的中值速率提升2.2倍;在UE数目为8时,Het-LDN与LDN相比,用户吞吐量之和的中值速率提升2.5倍;在UE数目为16时,Het-LDN

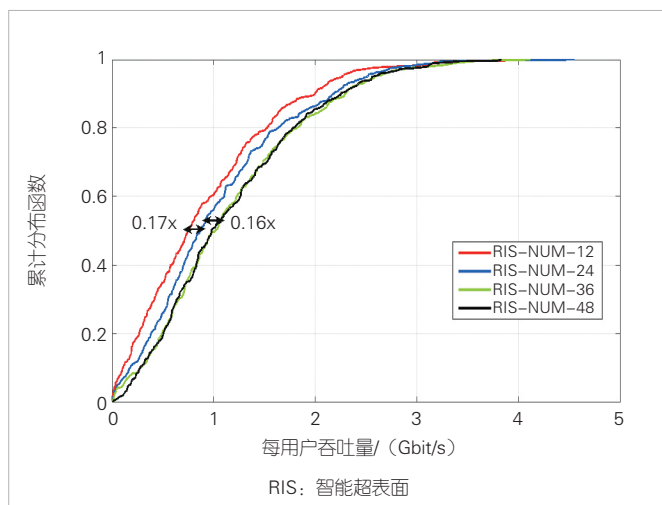


▲图6 不同UE数目下Het-LDN与LDN的所有用户吞吐量之和的累计分布函数(CDF)曲线

与LDN相比,用户吞吐量之和的中值速率提升4.7倍。此外,我们还可以看到,随着UE数目的增加,Het-LDN的吞吐量之和持续上升。这表明此系统可支持更多的UE进行同时传输,而LDN的吞吐量变化趋势则相反。上述的结果均表明Het-LDN的性能显著优于LDN。

2.3 不同参数配置下Het-LDN的性能评估

考虑到不同的参数配置会影响Het-LDN系统的性能。本小节中,我们主要评估RIS数目、UE数目和预编码方案对整体性能的影响。首先需要注意的是,在分析不同RIS数目对Het-LDN系统的影响时,不同仿真用例的仿真参数除了RIS数目外其他的参数均相同。于是,本部分的仿真中假设固定的AP数目为4,UE数目为4,RIS数目为12、24、36、48,其余的仿真参数见表3。由于系统采用RIS以等间隔分布的方式在四周墙壁上部署的策略,因此,当RIS数目增加时,每一面墙壁上RIS的增加数目按照RIS总体数目增长的比例而增加。图7给出了不同RIS数目下Het-LDN的每用户吞吐量的CDF曲线。可以发现,当RIS数目分别从12增加到24、从24增加到36时,中值用户的吞吐量分别提升17%和16%;而当RIS数目从36到48时,中值用户的吞吐量无明显变化。出现上述现象的原因是:当RIS数目小于36时,每一面墙壁上的RIS不够密集,则存在某些UE无法选择出到其他UE干扰较小的RIS的情况,因此,增加RIS的数目可减少上述情况的出现,从而保证系统获得更优的性能;而当RIS数目大于36时,每一面墙壁上的RIS密度足够,可以保证每个UE能选择出到其他UE干扰较小的RIS,系统整体性能达到饱和,此时



▲图7 不同RIS数目下Het-LDN的每用户吞吐量的累计分布函数(CDF)曲线

继续增加 RIS 数目, 对 Het-LDN 的整体性能提升作用将不明显。

此外, UE 数目和预编码方案也会影响 Het-LDN 的性能。本部分中不同的仿真用例除了 UE 数目的参数不同, 其他的参数均相同。其中, RIS 数目为 32, 其余的参数见表 3。图 8 为不同 UE 数目和不同预编码方案下 Het-LDN 的每用户吞吐量的 CDF 曲线。可以发现, 当采用 SVD 预编码方案且 UE 数目从 10 降低到 8 时, 中值用户的吞吐量提升 28%; 而当 UE 数目从 8 降低到 4 时, 中值用户的吞吐量则能提升 85%。产生上述现象的原因是: 随着可调度 UE 数目的减小, 多用户传输造成的用户间干扰降低, Het-LDN 系统的性能随之提升。当采用最大比传输 (MRT) [20] 预编码方案时, 可调度 UE 的数目与 Het-LDN 系统的性能之间也呈现出类似的规律。此外, 在 UE 数目分别是 10、8 和 4 的情况下, 比较 SVD 和 MRT 两种预编码对 Het-LDN 性能的影响时, SVD 预编码相对于 MRT 预编码, 其中值用户的吞吐量分别提升 17%、22% 和 15%。这是由于 MRT 预编码仅考虑增强有用接收信号的功率, 并没有考虑处理 MU 干扰, 而 SVD 预编码则考虑了有用信号和干扰信号之间的平衡, 因此, SVD 预编码的性能优于 MRT 预编码。

3 结束语

随着信息技术的不断发展, 各种智能设备的能力越来越强, 各类网络节点的部署越来越密, 服务质量的要求越来越高。因此面向未来的网络部署将面临着成本、能效、性能、复杂度和可扩展性等诸多挑战, 由多类不同功能的网络节点组成的 Het-LDN 可以有效应对上述的问题。两类不同功能

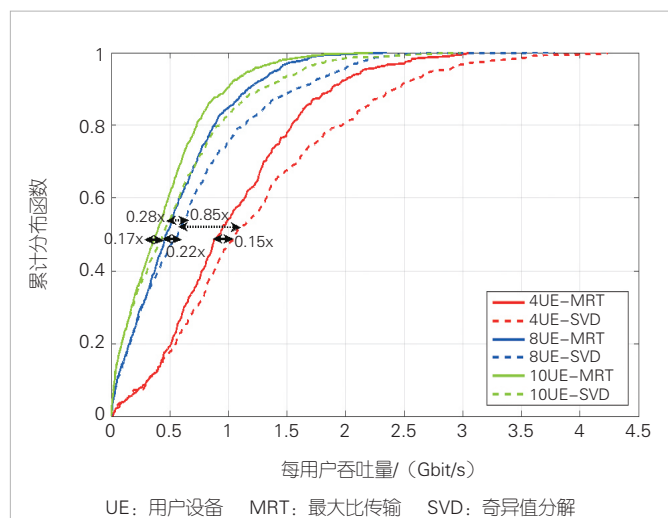
的网络节点有不同的候选方案, 需要综合考虑硬件成本、实现复杂度、功耗和可扩展性等多种不同因素确定最终节点的选型。本文中, 我们通过详细分析确定 Type-I-AP 和 Type-II-RN 分别优选 Option 7.2x AP 和 RIS。此外, 还通过系统级仿真实验进一步验证 Het-LDN 的性能。实验结果表明: Het-LDN 相对传统的 LDN 有着显著的性能优势; Het-LDN 中 RIS 数目对系统性能的影响存在限制, 当 RIS 数目增加到一定程度时, 继续增加 RIS 数目对系统性能提升没有贡献; Het-LDN 中使用 SVD 预编码性能优于 MRT 预编码。

致谢

感谢中兴通讯股份有限公司禹宏康、王云璐和张淑娟等对本研究的帮助!

参考文献

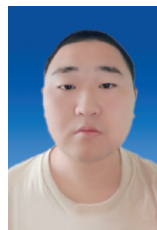
- [1] Ericsson. Radio stripes: re-thinking mobile networks (2019) [EB/OL]. [2024-01-12]. <https://www.ericsson.com/en/blog/2019/2/radio-stripes>
- [2] NGO H Q, ASHIKHMIN A, YANG H, et al. Cell-Free Massive MIMO: uniformly great service for everyone [C]//Proceedings of IEEE 16th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC). IEEE, 2015: 201-205. DOI: 10.1109/spawc.2015.7227028
- [3] INTERDONATO G, BJORSON E, NGO H Q, et al. Ubiquitous cell-free massive MIMO communications [EB/OL]. (2018-04-10) [2024-01-12]. <https://arxiv.org/abs/1804.03421>
- [4] BASHAR M, CUMANAN K, BURR A G, et al. Cell-free massive MIMO with limited backhaul [C]//Proceedings of IEEE International Conference on Communications (ICC). IEEE, 2018. DOI: 10.1109/icc.2018.8422865
- [5] BJORNSON E, SANGUINETTI L. Scalable cell-free massive MIMO systems [J]. IEEE transactions on communications, 2020, 68(7): 4247-4261. DOI: 10.1109/tcomm.2020.2987311
- [6] WEI C, XU K, XIA X C, et al. User-centric access point selection in cell-free massive MIMO systems: a game-theoretic approach [J]. IEEE communications letters, 2022, 26(9): 2225-2229. DOI: 10.1109/lcomm.2022.3186350
- [7] LARSSON E G, VIEIRA J. Phase calibration of distributed antenna arrays [J]. IEEE communications letters, 2023, 27(6): 1619-1623. DOI: 10.1109/lcomm.2023.3266836
- [8] KHANDEKAR A, BHUSHAN N, TINGFANG J, et al. LTE-advanced: Heterogeneous networks [C]//2010 European Wireless Conference (EW). IEEE: 978-982. DOI: 10.1109/EW.2010.5483516
- [9] HU R Q, QIAN Y. An energy efficient and spectrum efficient wireless heterogeneous network framework for 5G systems [J]. IEEE communications magazine, 2014, 52(5): 94-101. DOI: 10.1109/mcom.2014.6815898
- [10] 3GPP TR 38.801. Study on new radio access technology: Radio access architecture and interfaces (Release 14): 3GPP TR 38.801 [S]. 2017
- [11] O-RAN Alliance. Control, user and synchronization plane Specification: O-RANWG4.CUS.0-v10.00 [S]. 2022



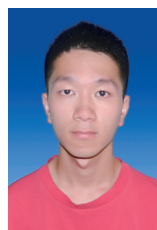
▲图 8 不同 UE 数目和不同预编码方案下 Het-LDN 的每用户吞吐量的累计分布函数(CDF)曲线

- [12] MIKIO I, HIDEAKI T, SATOSHI N. Relay technology in LTE-Advanced [J]. NTT DOCOMO technical journal, 2011, 12(2): 29-36
- [13] 3GPP. Study on integrated access and backhaul (Release 16): 3GPP TR 38.874 [S]. 2018
- [14] 3GPP. Study on NR network-controlled repeaters (Release 18): 3GPP [S]. 2022
- [15] ZHANG Z J, DAI L L. Reconfigurable intelligent surfaces for 6G: nine fundamental issues and one critical problem [J]. Tsinghua science and technology, 2023, 28(5): 929-939. DOI: 10.26599/tst.2023.9010001
- [16] 3GPP. Study on channel model for frequencies from 0.5 to 100 GHz (Release 17): 3GPP TR 38.901 [S]. 2022
- [17] ZHANG D, PAN P, YOU R N, et al. SVD-based low-complexity hybrid precoding for millimeter-wave MIMO systems [J]. IEEE communications letters, 2018, 22(10): 2176-2179. DOI: 10.1109/lcomm.2018.2865731
- [18] SRINIVASAN R. IEEE 802.16m evaluation methodology document: IEEE 802.16m-08/004r2 [R]. 2008
- [19] GU Q, WU D, SU X, et al. System-level Simulation of RIS assisted Wireless Communications System [C]//Proceedings of GLOBECOM 2022 - 2022 IEEE Global Communications Conference. IEEE, 2022: 1540-1545. DOI: 10.1109/globecom48099.2022.10000700
- [20] HAO L, ZHAO H, HU F L, et al. Energy efficient power allocation algorithm for downlink massive MIMO with MRT precoding [C]//Proceedings of IEEE 78th Vehicular Technology Conference (VTC Fall). IEEE, 2013: 1-5. DOI: 10.1109/vtcfall.2013.6692185

作者简介



叶新泉，中兴通信股份有限公司技术预研工程师；主要从事无线通信系统的物理层预研工作，主要研究方向为分布式多天线技术、可重构智能表面等；申请专利10余项。



卢光延，中兴通信股份有限公司技术预研工程师；主要从事无线通信系统的物理层预研工作，主要研究方向包括可重构智能表面、分布式多天线技术等；发表论文7篇。



陈艺骥，中兴通讯股份有限公司技术预研资深专家、青年领军人才、深圳市地方级领军人才；主要研究领域为6G技术预研；发表论文10篇，获授权专利50项。

下一代多址接入技术的挑战与关键进展



Challenges and Key Progress in Next Generation Multiple Access Technology

陈为/CHEN Wei^{1,2}, 艾渤/AI Bo^{1,2}

(1. 北京交通大学, 中国 北京 100044;

2. 先进轨道交通自主运行全国重点实验室, 中国 北京 100044)

(1. Beijing Jiaotong University, Beijing 100044, China;

2. State Key Laboratory of Advanced Rail Autonomous Operation, Beijing 100044, China)

DOI: 10.12142/ZTETJ.202401014

网络出版地址: <http://kns.cnki.net/kcms/detail/34.1228.TN.20240226.1537.011.html>

网络出版日期: 2024-02-27

收稿日期: 2023-12-20

摘要: 下一代多址接入技术将突破以正交性为基础的设计原则, 通过非正交性的设计和更先进的信息处理技术获得更高的谱效和能效。针对海量随机接入与非正交多址接入(NOMA)面临的不同挑战, 认为不同的接入方法与接收机设计可大幅提升大规模接入与数据传输的效率。此外, 人工智能、大规模多输入多输出(MIMO)和多基站协作等技术的蓬勃发展也为下一代多址技术的发展提供了新的可能。

关键词: 海量随机接入; NOMA; 人工智能; 大规模MIMO

Abstract: The next generation of multiple access technology will break through the design principles based on orthogonality, and achieve higher spectral and energy efficiency through non-orthogonality design and more advanced information processing technologies. In response to the different challenges faced by massive random access and non-orthogonal multiple access (NOMA), it is believed that different access methods and receiver designs can significantly improve the efficiency of large-scale access and data transmission. In addition, the flourishing development of technologies such as artificial intelligence, massive multiple input multiple output (MIMO), and multi-cell cooperation has also provided new possibilities for the development of next-generation multiple access technology.

Keywords: massive random access; NOMA; artificial intelligence; massive MIMO

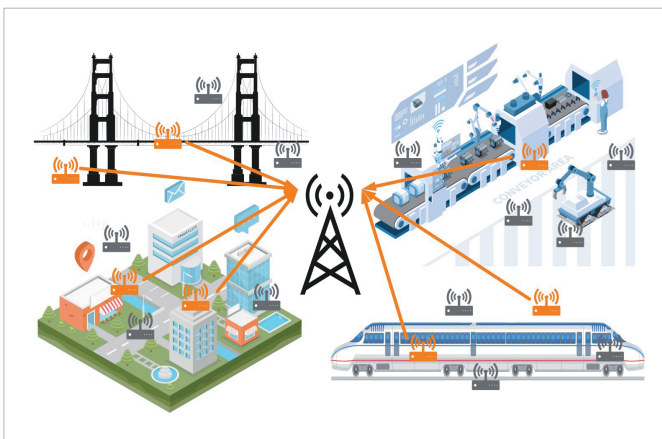
引用格式: 陈为, 艾渤. 下一代多址接入技术的挑战与关键进展 [J]. 中兴通讯技术, 2024, 30(1): 82-88. DOI: 10.12142/ZTETJ.202401014

Citation: CHEN W, AI B. Challenges and key progress in next generation multiple access technology [J]. ZTE technology journal, 2024, 30(1): 82-88. DOI: 10.12142/ZTETJ.202401014

1 发展下一代多址技术的背景与意义

为推动社会数字化转型, 物联网、智能城市、智能交通、远程医疗、智能工厂等各种应用蓬勃发展, 全球范围内的连接设备数量显著增长。面向如图1所示的各类新兴应用带来的海量连接, 我们需要为6G设计合适的多址技术, 即下一代多址技术, 这样能够让海量用户在给定的无线资源上高效且灵活地连接网络, 传输数据。传统的正交多址因其低效的资源分配策略和有限的支持用户数而难以满足6G通信系统提出的挑战性要求。因此, 下一代多址技术将突破以正交性为基础的设计原则, 通过非正交性的设计和更

先进的信息处理技术获得更高的谱效和能效。



▲图1 新兴应用下的海量连接示意图

基金项目: 国家自然科学基金(62122012、62221001); 中兴通讯产学研合作基金(HC-CN-20221208006)

1.1 海量随机接入的挑战

海量机器类型通信 (mMTC) 是 5G 的三大典型应用场景之一, 具有海量设备接入、零星通信、小数据包传输等特点。在此基础上, 6G 进一步提出了超大规模连接, 连接数高达 100 用户/平方米^[1], 其关键在于高效建立用户与基站的无线连接, 即海量用户的随机接入。在传统的基于授权的随机接入机制中, 频繁的交互导致信令开销比例过大, 小数据包传输的能效和谱效较低。免授权的随机接入机制无需等待基站调度即可进行信息传输, 尽管减小了信令开销, 却需要复杂的接收机算法实现活跃用户识别^[2]。因此, 如何设计适应海量连接特性的接入机制以实现万物智联成为一个关键挑战。此外, 为了在有限的频谱资源下支持更多用户接入, 非正交的资源分配被广泛应用。然而资源分配的非正交性导致用户间干扰, 接入的可靠性显著降低。如何在非正交资源分配下实现精确的活跃用户检测及信道估计成为重要挑战。

1.2 非正交多址接入的挑战

为了提高数据传输频谱效率和用户接入能力, 非正交多址接入 (NOMA) 允许不同用户复用无线资源, 并通过优化的资源分配策略和先进的接收机设计实现系统整体容量的提升。其中, 具有高谱效、大容量优势的功率域非正交多址接入 (PD-NOMA)^[3] 的研究较为广泛, 但仍有许多挑战需要应对。首先, 准确的信道状态信息 (CSI) 对于用户配对和连续干扰消除 (SIC) 至关重要。CSI 的估计错误可能导致用户分组、排序和数据解码错误。6G 通信系统中不断增加的连接用户数、更大规模的天线以及更高的用户移动性都将使 CSI 的获取更具挑战性。另外, 由于 NOMA 利用接收机的复杂度来换取高谱效, 如何在保证正确检测多用户信号的同时简化接收机的设计也是一个挑战性难题。

2 海量随机接入发展趋势

2.1 压缩感知随机接入

由于海量随机接入具备零星活跃的特性, 活跃用户检测及信道估计可以刻画成稀疏线性逆问题。利用经典的压缩感知算法, 如正交匹配追踪算法、迭代软阈值算法和近似消息传递算法等对稀疏信号进行恢复能够实现活跃用户检测与信道估计^[4]。在此基

础上, 可以利用星座点信息^[2]、活跃用户时间相关性^[5]以及信道空间相关性^[6]等附加信息进一步提高活跃用户检测与信道估计的准确性。然而, 压缩感知理论指出导频长度与活跃用户数成正比, 随着活跃用户数目增加, 用户所需的频谱资源不断增长, 因此如何在有限的频谱资源下支持尽可能多的用户接入是压缩感知随机接入需要解决的重要问题。此外, 已有算法的计算复杂度较高, 如何设计高效且低复杂度的算法以实现活跃用户检测和信道估计也是一个挑战性难题^[4]。

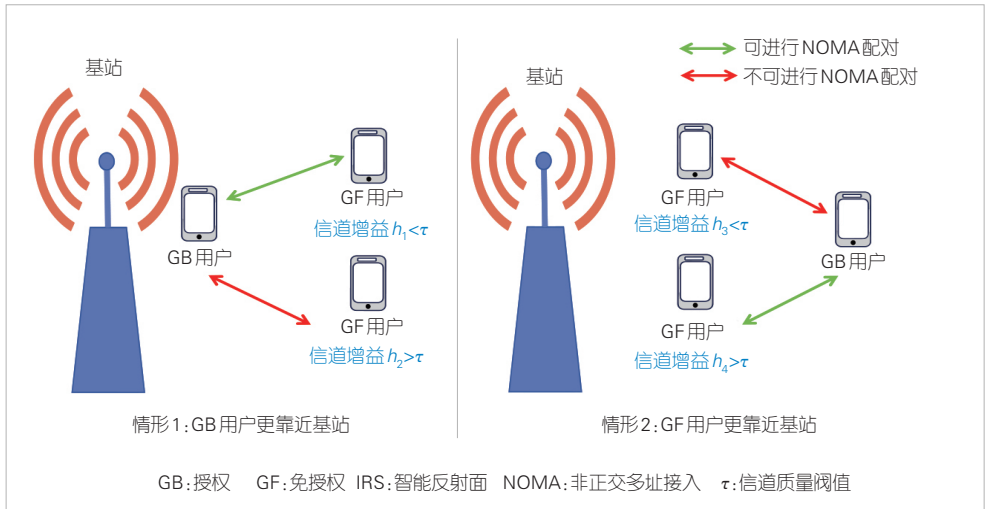
2.2 半免授权随机接入

在授权接入方式下, 用户可获得基站分配的专用频谱资源从而建立稳定通信连接, 接入可靠性高。免授权接入方式下, 用户无须等待基站调度即可进行信息传输, 接入时延较低。然而, 授权接入与免授权接入在面对大规模连接时均具有一定局限性。表 1 给出了两种接入方式的对比。

在实际通信场景中, 针对各类用户对接入可靠性和时延的不同需求, 融合授权和免授权接入的半免授权随机接入成为重要的研究方向。其中, NOMA 辅助的半免授权随机接入是研究热点。NOMA 允许免授权用户复用授权用户的频谱资源, 并通过信道质量阈值控制参与资源复用的免授权用户数

▼表1 授权接入与免授权接入的对比

接入方式	握手协议	优点	缺点
授权接入	4步/2步	连接稳定	信令开销和接入延迟较大
免授权接入	无需上行调度请求	低信令开销与接入时延	易发生频繁的用户碰撞, 接入性能不稳定



▲图2 IRS辅助的大规模多输入多输出-非正交多址(MIMO-NOMA)用户配对

目以降低其对授权用户造成的性能损失。根据授权用户与免授权用户空间分布的不同,图2给出了两种智能反射面(IRS)辅助的半免授权接入场景。为保证用户服务质量,复用同一频谱资源的用户需满足不同信道条件^[7]:授权用户靠近基站而免授权用户是蜂窝边缘用户,为增大授权与免授权用户之间的信道增益差,基站向所有免授权用户广播信道质量阈值,信道增益小于该阈值的免授权用户才能与授权用户进行配对;授权用户是蜂窝边缘用户而免授权用户靠近基站,信道增益大于信道质量阈值的免授权用户与授权用户进行配对。

2.3 无源随机接入

随机接入可分为有源和无源两种形式。在有源随机接入中,基站对传输的消息和生成信息的用户身份都感兴趣,用户将身份标识连同有效负载一起传输至基站以进行用户检测和数据恢复。

在mMTC场景下,为海量低活跃度用户分配不同身份标识会造成极大的资源开销。在无源随机接入中,基站只对接收到的信息内容感兴趣,而不关注信息的来源用户。该接入方式可视作一种特殊的免授权接入,每个活跃用户并不分配唯一前导序列而是共享公共码本,基站的任务是在不考虑接入用户身份信息的情况下正确解码所有接收数据^[8]。公共码本的使用能够节省用户接入时的前导开销,提升系统可容纳的设备总量。随着用户发送的载荷数据长度增长,公共码本下的非相干检测算法复杂度随之增加^[9]。为了降低接收端计算成本,可将接入用户数据拆成更短的分组进行传输。发送端添加分组顺序信息,解码器对分组数据及其排列顺序进行恢复,最终将同一用户的各个分组进行拼接获得原始信息^[10]。

2.4 AI赋能

人工智能(AI)可以基于用户的位置信息、接入历史和通信行为等建立用户画像,进而实现对用户接入请求的预测。对于在未来一段时间内发起接入请求概率较高的用户,AI可提前分配适量的通信资源,从而提高资源利用率并降低用户的接入失败率。此外,由于传统的活跃用户检测和信道估计算法可能存在问题建模误差和求解误差,且计算复杂度较高,基于AI的方法可将多用户检测问题视作端到端的映射问题。通过将传统算法展开为神经网络^[11],构建具有可训练参数的网络模型,并利用训练数据学习优化参数,AI可实现比传统多用户检测算法更优的检测性能,同时提升算法收敛速度和降低时延。值得注意的是,为了提升网络性

能,通常需要大量数据训练模型。基站中积累的真实用户接入数据可作为数据来源之一。当真实数据较少时,可考虑使用生成对抗网络或其他数据生成模型得到模拟用户接入数据,基于该数据进行模型训练后再利用真实数据对模型进行微调。

由于不同接入场景的信噪比和活跃用户数等不尽相同,多用户检测模型的泛化性是重要的性能评价指标。为减少面向不同场景学习特定网络模型所带来的极大存储和训练开销,灵活地适应时变的通信场景,元学习仅通过使用少量的新训练数据即可快速适应新的应用场景^[12]。网络的复杂度也同样影响系统性能。相比于低复杂度模型,高复杂度模型可更准确地拟合用户接入数据,但也可能出现过拟合以及存储开销大的问题。采用模型剪枝等技术对完成训练的网络进行压缩可减少冗余参数,从而节省存储空间并简化计算^[13]。除此之外,神经网络的鲁棒性能够衡量网络对输入数据中的扰动、噪声及异常情况的抵抗能力,是评价网络的又一重要指标。考虑到多用户检测模型受到不同类型噪声的干扰可能出现性能下降,采用对抗防御方法增强模型鲁棒性可进一步提升多用户检测性能^[14]。

2.5 MIMO赋能

基站天线通常位于高建筑物上,此时远离基站且周围存在丰富散射点的用户信号仅具有微小角扩展,因此可将用户信号到达角作为区分不同活跃用户的依据,从而有效提升接入用户数^[15]。大规模多输入多输出(MIMO)的另一特性是存在信道硬化现象,即当基站天线数目很大时,用户与基站之间的随机信道特性趋于稳定,用户信道增益接近其平均信道增益。利用信道硬化现象,多天线基站无须估计用户信道的瞬时状态信息,仅基于用户的统计信道信息即可实现活跃用户检测^[16]。和传统基于压缩感知的用户检测算法相比,该方法在相同导频长度下可检测更多活跃用户,从而支持更大规模的用户接入。

2.6 多基站协作

在传统的非合作多基站系统中,每个基站独立检测来自本小区的用户,小区间干扰被视作有害噪声。然而,由于高频信号的高衰减特性以及用户对通信容量需求的不断增长,密集基站布局逐渐形成,同频小区之间的干扰愈发严重。为不同小区用户分配正交导频在抑制干扰的同时也会带来成倍的资源开销,而使用非正交导频使得用户检测算法性能受限于非正交的强度。

多基站协作技术通过基站之间的信息交互、资源共享和

联合优化,可显著降低小区间干扰或将干扰转化成有用信息。基站协作可通过集中或边缘处理两种方式来实现。在集中方式中,中央单元将接收到的用户信息进行集中处理,与其连接的基站跨多个小区合作以对有源设备进行检测,从而有效避免小区间干扰带来的检测性能下降,同时节省使用正交导频池所带来的额外资源开销^[17]。在边缘处理方式中无需中央单元,距离邻近的基站通过用户状态信息的交换即可实现准确的活跃用户检测^[18]。

3 NOMA 发展趋势

3.1 高移动性场景

PD-NOMA 作为 NOMA 的一个重要分支,允许用户共享同一频谱资源并在功率域区分用户,能够显著提升系统容量。在 PD-NOMA 中,合理的功率分配是有效实现 SIC 的关键。然而,在高移动性用户的时变信道下使用静态功率分配方案可能导致 SIC 失败,从而使得系统性能急剧下降。

1) NOMA-OTFS

正交时频空(OTFS)调制技术引入了延迟多普勒域信道表示和信号调制方法,能够支持高移动性场景下的可靠通信^[19]。NOMA-OTFS 根据用户移动性将一个高速移动用户和多个低速移动用户分为一组,通过 OTFS 在延迟多普勒平面和时频面区分不同移动性用户,再通过 NOMA 恢复多个低速移动用户数据,提高频谱资源利用率^[20]。

2) 用户分组

在 PD-NOMA 中,由于高移动性用户的信道条件不断变化,同一组内用户间的信道增益差距难以保证,从而导致 SIC 失败。我们可通过重组用户来调整用户间信道增益差从而解决该问题。例如,采用解离和重关联机制对用户间的瞬时信道增益差进行监测,当某一用户与其他用户的信道增益差小于给定阈值时,将自动识别分离该用户并将其重新关联到其他用户组中^[21]。

3.2 异步 NOMA

相较于同步 NOMA,在异步 NOMA 传输中不需要实现严格同步。通过合适的收发机设计,异步传输具有一些有益的特性。例如,异步传输被证明具有更高的多用户接入容量上界、更大的分集增益和更高的传输速率^[22]。上行异步 NOMA 可通过给不同用户的发送符号专门设置不匹配的时间来实现,下行异步 NOMA 可在发送端实现叠加编码之前,给不同用户的符号添加子符号延迟来实现^[4]。

3.3 AI 辅助的 NOMA

将 AI 与 NOMA 结合,能利用其在智能化辅助和决策支持等方面的优势,进一步提升 NOMA 通信系统的服务质量。

1) 功率分配

不同的用户功率分配方案会影响用户信号的解调效果。通过 AI 技术可以解决复杂场景下的用户关联、子信道和功率分配难题,提升系统性能^[23]。例如,子信道和功率分配的设计可分别采用半监督学习和深度神经网络(DNN)。此外,搜索最佳 NOMA 功率分配方案可利用包含多个卷积层和多个隐藏层的 DNN,以实现更高的数据速率和能量效率^[24]。

2) 用户检测

NOMA 系统可采用深度多任务学习将叠加编码传输视为多个不同但又相互关联的学习任务,建立统一的 NOMA 多任务 DNN 框架用于用户检测,实现端到端的系统性能优化^[25]。该方案与常规的 NOMA 方案相比,计算复杂度更低,且能在加性白高斯噪声信道、瑞利衰落信道等信道模型下实现更高的传输性能。

将 AI 技术与 NOMA 结合同样需要考虑模型的泛化性与鲁棒性等问题,其与 2.4 部分内容相似,此处不再赘述。

3.4 大规模 MIMO-NOMA 与去蜂窝大规模 MIMO-NOMA

3.4.1 大规模 MIMO-NOMA

大规模 MIMO 利用空间资源提高频谱效率,可与 NOMA 结合带来更多的性能增益。

1) 导频污染消除

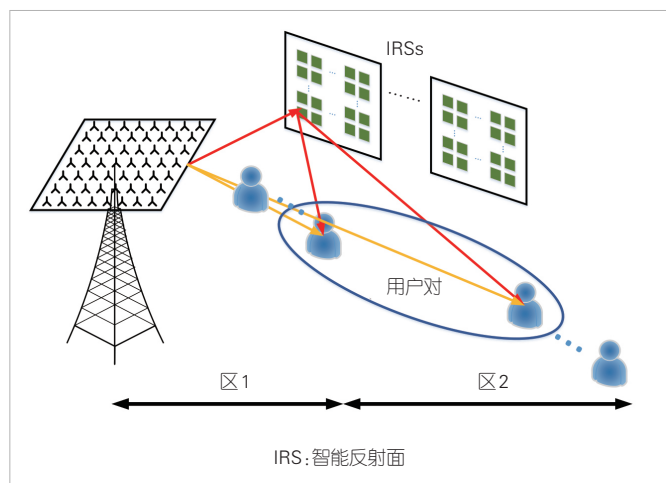
大规模 MIMO 中天线和用户数量庞大,而正交导频序列数量有限,为不同用户和天线分配非正交导频会导致导频污染。通过将每个小区中的用户分成多个组,不同组共用相同的正交导频集,并采用基于组间 SIC 的半盲信道估计方法可减少大规模 MIMO-NOMA 中同一小区内的组间导频污染^[26]。

2) IRS 辅助的用户配对

当大规模 MIMO-NOMA 中的用户具有不同的信道条件时,对用户进行合理配对是实现高传输速率的关键。如图 3 所示,在智能反射面(IRS)辅助的大规模 MIMO-NOMA 中,可利用用户相对于 IRS 的位置及其信道增益来对用户进行配对,以此设计自适应的用户配对算法,提升系统性能^[27]。

3.4.2 去蜂窝大规模 MIMO-NOMA

去蜂窝大规模 MIMO 作为改进式的分布式大规模 MIMO,能为用户提供较高的分集增益,进而提升系统吞吐量。



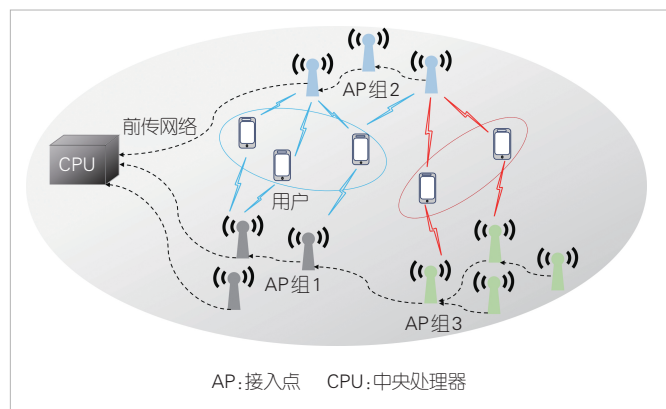
▲图3 IRS辅助的大规模多输入多输出-非正交多址用户配对

1) 用户分组

对去蜂窝大规模MIMO-NOMA网络中的用户进行分组，使每组中的用户具有尽可能大的信道增益差，可获取更好的系统性能。然而，去蜂窝大规模MIMO-NOMA系统中的每个用户通过同时接入多个接入点（AP）实现在多个信道上传输数据，如何衡量用户间的信道增益差以实现用户分组也是一大挑战。可根据网络中每个用户与所有AP之间的信道增益计算信道增益质心，利用该质心定义每个用户的杰卡德相似系数。通过用户间的杰卡德相似系数的差值来衡量两个用户信道增益差距，以此对用户进行分组，能提升系统数据传输速率^[28]。

2) 接入点聚类

在去蜂窝大规模MIMO-NOMA中，通过对AP聚类来优化波束形成矢量，能够降低系统信号检测的计算复杂度及处理延迟，实现系统性能的提升。如图4所示的动态聚类去蜂窝上行传输网络，可根据网络覆盖范围内的AP与用户间的



▲图4 动态聚类去蜂窝上行传输网络

CSI进行AP聚类，每组AP作为一个独立的分布式天线系统为网络覆盖范围内的所有用户提供服务。通过共同优化AP聚类和波束成形来最大化总用户传输速率，能获取更佳的系统性能^[29]。

3.5 其他类型的NOMA技术

除上述提到的PD-NOMA以外，还有多种其他类型的NOMA技术，如码域NOMA中的稀疏码多址（SCMA）、多用户共享接入（MUSA）以及多域NOMA中的图样分割多址（PDMA）等。

3.5.1 SCMA

SCMA的思想主要包括以下几个方面：1) 将二进制数据比特编码为预先设计的码本中的多维稀疏码字；2) 为不同用户分配不同码本以实现多址；3) 利用码字的稀疏性，接收端可采用消息传递算法（MPA）对用户信号解码^[30]。与CD-NOMA中的其他技术相比，采用多维星座码本的SCMA能获取更高性能增益，且具有更低误码的率。然而，SCMA也存在一些挑战。首先，由于SCMA将数据比特映射为具有多维星座的稀疏码字进行传输，因此多维星座的设计方案将密切影响系统性能。另外，大部分解码用户信号的MPA复杂度高，计算开销大。在现有方法中，可通过最大化用户星座点间的欧氏距离改进星座设计方案，也可通过改进MPA算法、限制码字搜索空间、结合深度学习等方式优化SCMA中的用户信号解码^[31]。

3.5.2 MUSA

MUSA可看作一种基于扩展序列的码域NOMA方案，能够在低控制开销和功耗下实现高过载率、低延迟的海量用户接入，适用于mMTC等大规模接入场景^[32]。MUSA的核心思想是非正交传输和免授权接入。多个用户的数据信息经过编码、调制、扩频后，可在同一频谱资源上进行免授权传输，并在接收端使用SIC对信号进行解码。其中，扩频序列的设计直接影响MUSA的过载性能以及接收机的复杂度，是实现MUSA的关键步骤。MUSA的扩频序列被设计为短长度的复扩展序列。序列的实部和虚部设计相对自由以确保扩展码的长度不会过长，从而有效降低系统功耗和延迟并增强用户过载能力。同时，特殊设计的扩频序列之间具有低相关性，可有效减小接收端复杂度。MUSA还可利用多用户与基站间的远近效应带来的接收信号信噪比差异来增强SIC。因此，MUSA不需要精确的闭环功率控制，进一步降低了系统的控制开销。

然而,系统过载比过高时,基于压缩感知的多用户检测算法由于感知矩阵列相关性增加而出现性能下降。因此,为MUSA开发实际可行的、可扩展的多用户检测方案是一个具有挑战性的开放问题^[33]。

3.5.3 PDMA

基于发射机和接收机联合设计的PDMA是一种极具潜力的NOMA技术^[34]。PDMA编码图样定义了用户数据到物理资源的映射关系,是该接入方式的核心内容。PDMA将系统图样表示为0/1矩阵形式,每行对应系统的可用资源单元,每列对应系统的活跃用户。矩阵元素为“1”时表示对应用户数据映射到相应资源。不同用户对系统可用资源的不同占用方式形成用户特有的编码图样。在发送端,多个用户的待传输符号通过各自的编码图样被映射到同一物理资源(时域、频域和空间域等资源)上进行叠加传输,接收端采用非线性检测算法进行信号解码。PDMA系统允许不同用户的编码图样存在部分重叠,但仍然可以通过未碰撞的分量实现用户区分。因此,在不增加物理资源的前提下,通过将用户对物理资源的占用方式作为资源共享的新维度,PDMA系统可减少因用户碰撞导致的接入失败。

矩阵每一列中“1”的个数可视作用户信号的分集阶数。更大的分集阶数可以实现更可靠的数据传输,但也会使接收端信号处理复杂度增加。在计算资源有限时,可使用稀疏的PDMA图样,即每位用户的信号只占用少部分的可用资源,从而在保证系统可靠性的同时降低接收算法复杂度^[34]。PDMA的另一优势在于接收机可通过数字信号处理器实现并行干扰消除。相比于易受到误差传播影响的SIC算法,采用并行干扰消除算法(如MPA)能够进一步提高稀疏模式编码的PDMA解码正确率^[35]。

4 结束语

面向6G万物智联的新兴业务和更高需求,下一代多址技术是实现海量用户随机接入和高谱效/能效数据传输的关键。压缩感知、半免授权接入及无源接入等新型接入方法和NOMA为大规模随机接入和数据传输提供了新的研究思路。人工智能等新技术有望与下一代多址技术深度融合,提升随机接入与数据传输的效率,降低传输时延,简化接收机设计,在未来6G各类应用场景中发挥重要作用。

参考文献

[1] WANG C X, YOU X H, GAO X Q, et al. On the road to 6G: visions, requirements, key technologies, and testbeds [J]. IEEE

communications surveys & tutorials, 2023, 25(2): 905–974. DOI: 10.1109/COMST.2023.3249835

[2] BAI Y N, CHEN W, AI B, et al. Prior information aided deep learning method for grant-free NOMA in mMTC [J]. IEEE journal on selected areas in communications, 2022, 40(1): 112–126. DOI: 10.1109/JSAC.2021.3126071

[3] CHAUHAN A, GHOSH S, JAISWAL A. RIS partition-assisted non-orthogonal multiple access (NOMA) and quadrature-NOMA with imperfect SIC [J]. IEEE transactions on wireless communications, 2023, 22(7): 4371–4386. DOI: 10.1109/TWC.2022.3224645

[4] CHEN W, LIU Y W, JAFARKHANI H, et al. Signal processing and learning for next generation multiple access in 6G [EB/OL]. (2023–09–01)[2024–01–22]. <http://export.arxiv.org/abs/2309.00559v2>

[5] WANG Q P, LIU L, ZHANG S W, et al. Exploiting temporal side information in massive IoT connectivity [J]. IEEE transactions on wireless communications, 2023, 22(2): 1432–1447. DOI: 10.1109/twc.2022.3204768

[6] DJELOUAT H, LEINONEN M, RIBEIRO L, et al. Joint user identification and channel estimation via exploiting spatial channel covariance in mMTC [J]. IEEE wireless communications letters, 2021, 10(4): 887–891. DOI: 10.1109/lwc.2021.3049167

[7] DING Z G, SCHÖBER R, POOR H V. A new QoS-guarantee strategy for NOMA assisted semi-grant-free transmission [J]. IEEE transactions on communications, 2021, 69(11): 7489–7503. DOI: 10.1109/tcomm.2021.3100598

[8] AHMADI M J, DUMAN T M. Random spreading for unsourced MAC with power diversity [J]. IEEE communications letters, 2021, 25(12): 3995–3999. DOI: 10.1109/lcomm.2021.3119424

[9] KE M L, GAO Z, ZHOU M Y, et al. Next-generation URLLC with massive devices: a unified semi-blind detection framework for sourced and unsourced random access [J]. IEEE journal on selected areas in communications, 2023, 41(7): 2223–2244. DOI: 10.1109/JSAC.2023.3280981

[10] AMALLADINNE V K, CHAMBERLAND J F, NARAYANAN K R. A coded compressed sensing scheme for unsourced multiple access [J]. IEEE transactions on information theory, 2020, 66(10): 6509–6533. DOI: 10.1109/tit.2020.3012948

[11] LI Y, CHEN Z L, WANG Y Q, et al. Heterogeneous transformer: a scale adaptable neural network architecture for device activity detection [J]. IEEE transactions on wireless communications, 2023, 22(5): 3432–3446. DOI: 10.1109/TWC.2022.3218579

[12] HUO H, XU J, SU G, et al. Intelligent MIMO detection using meta learning [J]. IEEE Wireless Communications Letters, 2022, 11(10): 2205–2209. DOI: 10.1109/LWC.2022.3197158

[13] RUI L, YANG S, CHEN S, et al. Smart network maintenance in an edge cloud computing environment: an adaptive model compression algorithm based on model pruning and model clustering [J]. IEEE transactions on network and service management, 2022, 19(4): 4165–417. DOI: 10.1109/TNSM.2022.3202796

[14] LIU A, LIU X, YU H, et al. Training robust deep neural networks via adversarial noise propagation [J]. IEEE transactions on image processing, 2021, 30: 5769–5781

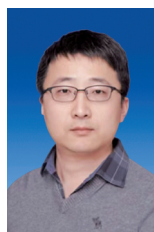
[15] CHEN W, XIAO H, SUN L, et al. Joint activity detection and channel estimation in massive MIMO systems with angular domain enhancement [J]. IEEE transactions on wireless communications, 2022, 21(5): 2999–3011. DOI: 10.1109/TWC.2021.3117358

[16] CHEN Z, SOHRABI F, LIU Y F, et al. Phase transition analysis for covariance-based massive random access with massive MIMO [J]. IEEE transactions on information theory, 2021, 68(3): 1696–1715

[17] CHEN Z L, SOHRABI F, YU W. Sparse activity detection in multi-cell massive MIMO exploiting channel large-scale fading [J].

- IEEE transactions on signal processing, 2021, 69: 3768–3781. DOI: 10.1109/TSP.2021.3090679
- [18] SHAO X D, CHEN X M, NG D W K, et al. Cooperative activity detection: sourced and unsourced massive random access paradigms [J]. IEEE transactions on signal processing, 2020, 68: 6578–6593. DOI: 10.1109/TSP.2020.3039342
- [19] 刘梦晓, 周昆, 张文逸. 用于超高移动性信道的正交时频空调制 [J]. 中兴通讯技术, 2021, 27(4): 9–13. DOI: 10.12142/ZTETJ.202104003
- [20] DING Z G, SCHÖBER R, FAN P Z, et al. OTFS-NOMA: an efficient approach for exploiting heterogeneous user mobility profiles [J]. IEEE transactions on communications, 2019, 67(11): 7950–7965. DOI: 10.1109/tcomm.2019.2932934
- [21] NAEEM M K, ABOZARIBA R, ASADUZZAMAN M, et al. Mobility support for MIMO-NOMA user clustering in next-generation wireless networks [J]. IEEE transactions on mobile computing, 2023, 22(10): 6011–6026. DOI: 10.1109/TMC.2022.3186430
- [22] GANJI M, ZOU X, JAFARKHANI H. Asynchronous transmission for multiple access channels: rate-region analysis and system design for uplink NOMA [J]. IEEE transactions on wireless communications, 2021, 20(7): 4364–4378. DOI: 10.1109/twc.2021.3058405
- [23] ZHANG H J, ZHANG H S, LONG K P, et al. Deep learning based radio resource management in NOMA networks: user association, subchannel and power allocation [J]. IEEE transactions on network science and engineering, 2020, 7(4): 2406–2415. DOI: 10.1109/tnse.2020.3004333
- [24] HUANG H, YANG Y, DING Z, et al. Deep learning-based sum data rate and energy efficiency optimization for MIMO-NOMA systems [J]. IEEE transactions on wireless communications, 2020, 19(8): 5373–5388
- [25] YE N, LI X M, YU H X, et al. DeepNOMA: a unified framework for NOMA using deep multi-task learning [J]. IEEE transactions on wireless communications, 2020, 19(4): 2208–2225. DOI: 10.1109/twc.2019.2963185
- [26] HU C, WANG H, SONG R F. Group successive interference cancellation assisted semi-blind channel estimation in multi-cell massive MIMO-NOMA systems [J]. IEEE communications letters, 2021, 25(9): 3085–3089. DOI: 10.1109/lcomm.2021.3095119
- [27] PERDANA R H Y, NGUYEN T V, AN B. Adaptive user pairing in multi-IRS-aided massive MIMO-NOMA networks: spectral efficiency maximization and deep learning design [J]. IEEE transactions on communications, 2023, 71(7): 4377–4390. DOI: 10.1109/tcomm.2023.3277533
- [28] REZAEI F, HEIDARPOUR A R, TELLAMBURA C, et al. Underlaid spectrum sharing for cell-free massive MIMO-NOMA [J]. IEEE communications letters, 2020, 24(4): 907–911. DOI: 10.1109/lcomm.2020.2966195
- [29] AL-ERYANI Y, AKROUT M, HOSSAIN E. Multiple access in cell-free networks: outage performance, dynamic clustering, and deep reinforcement learning-based design [J]. IEEE journal on selected areas in communications, 2021, 39(4): 1028–1042. DOI: 10.1109/jsac.2020.3018825
- [30] NIKOPOUR H, BALIGH H. Sparse code multiple access [C]// Proceedings of IEEE 24th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC). IEEE, 2013: 332–336. DOI: 10.1109/pimrc.2013.6666156
- [31] REBHI M, HASSAN K, RAOOF K, et al. Sparse code multiple access: potentials and challenges [J]. IEEE open journal of the communications society, 2021, 2: 1205–1238. DOI: 10.1109/ojcoms.2021.3081166
- [32] YUAN Z F, YU G H, LI W M, et al. Multi-user shared access for Internet of Things [C]// Proceedings of IEEE 83rd Vehicular Technology Conference (VTC Spring). IEEE, 2016: 1–5. DOI: 10.1109/vtcspring.2016.7504361
- [33] THUSHAN S, ALI S, MAHMOOD N H, et al. Deep learning-based blind multiple user detection for grant-free SCMA and MUSA systems [J]. IEEE transactions on machine learning in communications and networking, 2023, 1: 61–77. DOI: 10.1109/tmlcn.2023.3283350
- [34] DAI X M, ZHANG Z Y, BAI B M, et al. Pattern division multiple access: a new multiple access technology for 5G [J]. IEEE wireless communications, 2018, 25(2): 54–60. DOI: 10.1109/mwc.2018.1700084
- [35] SHEN S, CHEN Y W, ZHOU Q, et al. Demonstration of pattern division multiple access with message passing algorithm for multi-channel mmWave uplinks via RoF mobile fronthaul [J]. Journal of lightwave technology, 2020, 38(21): 5908–5915. DOI: 10.1109/JLT.2020.3005905

作者简介



陈为, 北京交通大学教授; 长期从事无线通信、信号处理、人工智能技术研究, 并围绕高维信息感知和处理、海量机器类通信、语义通信、智慧交通开展研究工作; 作为负责人承担国家优秀青年科学基金等多项省部级及以上项目; 获2022年度信息通信领域十大科技进展、首届詹天佑铁道科学技术奖秦驰道专项奖(青年奖)等荣誉。



艾渤, 北京交通大学教授、电子信息工程学院院长, 国家6G技术研发总体专家组专家, 国家自然科学基金委创新群体带头人, 国家杰出青年科学基金、优秀青年科学基金、牛顿高级学者基金、中国科协“求是杰出青年奖”获得者, 中共中央组织部“万人计划”领军人才, 北京市优秀教师, IEEE Fellow, IEEE VTS杰出讲师; 入选斯坦福2020—2022年全球前2%顶尖科学家终身影响力榜, 获省部级科技奖励9项。

数据中心光模块技术及演进



Data Center Optical Module Technology and Its Evolution

张平化/ZHANG Pinghua, 王会涛/WANG Huitao,
付志明/FU Zhiming

(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.202401015

网络出版地址: <http://kns.cnki.net/kcms/detail/34.1228.TN.20240304.1751.002.html>

网络出版日期: 2024-02-28

收稿日期: 2024-01-16

摘要: 算力需求提升带动网络带宽成倍增加, 数据中心能耗呈指数型增长。针对低功耗、高带宽的技术需求, 硅光、共封装光学 (CPO) 等技术有望成为长期解决方案。硅光芯片采用光互联, 叠加 CPO 技术, 将光引擎与交换芯片共同封装, 在速率提高的同时大大缩减功耗。液冷技术、线性驱动可插拔光模块 (LPO)、相干技术及薄膜铌酸锂等技术成为光模块优化主要新趋势。LPO 在高线性度跨阻放大器 (TIA)/驱动芯片厂商大力推动下可快速落地。相干精简版解决方案在数据中心 2 km 以内传输距离方面有竞争优势。

关键词: 硅光; CPO; LPO; 薄膜铌酸锂

Abstract: Network bandwidth multiplies and power consumption of data centers increases exponentially. SiPh and co-package optics (CPO) may be long-term solutions around the technical requirements of low-power consumption and high-bandwidth. SiP chips can be optical interconnection. Using SiPh and CPO stacking techniques which can realize co-package of optical engines and switch chips, power consumption can be greatly reduced when transmission rate increases. Besides, liquid cooling technology, linear-drive pluggable optics (LPO), coherent technology, and thin film lithium niobate will be main development trends of optical modules. LPO should be quick landing with vigorous promotion of high-linearity trans-impedance amplifier (TIA)/driver manufacturers. Coherent lite solution becomes competitive when solving under 2 km transmission in data centers.

Keywords: SiPh; CPO; LPO; thin film lithium niobate

引用格式: 张平化, 王会涛, 付志明. 数据中心光模块技术及演进 [J]. 中兴通讯技术, 2024, 30(1): 89-98. DOI: 10.12142/ZTETJ.202401015

Citation: ZHANG P H, WANG H T, FU Z M. Data center optical module technology and its evolution [J]. ZTE technology journal, 2024, 30(1): 89-98. DOI: 10.12142/ZTETJ.202401015

1 数据中心光模块技术创新点

作为数据存储和计算的中心, 数据中心在各个时期承载的主要功能有所差异。自 2000 年以来, 全球数据中心从计算中心发展到信息中心, 再发展至云计算中心, 目前正向算力中心演变。

1.1 算力时代下低功耗成为技术创新点

算力需求提升带动网络带宽成倍增加, 数据中心能耗呈指数型增长。设备厂商 Cisco 的数据显示^[1], 2010—2022 年全球数据中心的网络交换带宽提升了 80 倍, 背后的代价是交换芯片功耗增加约 8 倍, 光模块功耗增加 26 倍, 交换芯片串行器/解串器 (SerDes) 功耗增加 25 倍。

如何解决功耗问题成为下一代高速光互联应用的最大挑战。液冷技术成为数据中心突破节能瓶颈最有效的方式。液冷光模块是未来技术的一个亮点。

在光模块降耗的发展趋势下, 行业围绕驱动器、调制

器、激光器及电接口 4 个方面降低功耗。在驱动器方面, 如果采用线性驱动可插拔光模块 (LPO) 技术路线, 则可在数据链路中只使用线性模拟元件, 无需数字信号处理 (DSP) 或者时钟数据恢复 (CDR) 芯片, 从而降低 DSP/CDR 产生的功耗。在调制器方面, 薄膜铌酸锂基于自身较低的半波电压, 所需要的驱动电压更小, 从而具备优越的低功耗特性, 成为产业界尝试解决调制器功耗问题的重要途径。在激光器方面, 提高激光器的电光效率和耦合效率是两种降低功耗的途径。在电接口方面, 缩短交换芯片和光引擎之间的连接长度, 采用诸如近封装光学 (NPO) 和共封装光学 (CPO) 之类的超密集封装, 可实现更高密度的高速端口, 提升整机的带宽密度。而 NPO/CPO 背后的技术, 即硅光技术, 是以光子和电子为信息载体的硅基光电子大规模集成技术。

1.2 算力时代下高带宽需求是技术原动力

数据中心设备之间的连接由高速光接口提供, 并且根据

连接距离不同分为短距离（SR）、DR（指 500 m 距离）、远距离（FR）、长距离（LR）等规格，不同传输距离采用的技术方案也会有所不同。高速光接口的速率发展与数据中心的交换机容量以及 SerDes 技术的发展息息相关。交换机容量每 2 年翻一番，预计 2030 年会出现 400 Tbit/s 交换容量，单端口速率需要增长到 3.2 Tbit/s。

根据接收技术的不同，光连接技术可以分为直检检测技术和相干检测技术。直检检测技术由于成本低、功耗低，在 800GE 之前，为数据中心高速光接口的主要技术。随着速率的提升，直检检测技术受到色散、四波混频等问题的影响，传输距离下降，使得相干技术下沉到数据中心成为可能。在 800GE 时代，IEEE 802.3dj 针对 10 km 场景将定义相干和直检两条技术路径。但相干技术面临功耗高、成本高的挑战。未来 3.2T 时代，直检技术和相干技术将同时存在。直检检测技术在 3.2T 时代仍是主力技术路径之一。在单通道速率持续提升的同时，通过增加光纤或者波分复用技术来增加并行路数的方法也将持续发展。800GE 时代不仅使用单通道 100G 技术，还使用了单通道 200G 的技术。3.2T 时代将会依托单通道 200G 技术进行多路复用，当然也可能会发展单通道 400G 的技术。

2 光模块新技术演进路径

算力时代对光模块有着低功耗、高带宽的需求。目前业界技术演进方向如下：

1) 液冷光模块。冷板式液冷系统中的冷却液不与发热电子部件直接接触，所以目前常规光模块在冷板式液冷系统中应用时，一般无须考虑兼容性问题。浸没式和喷淋式液冷系统的冷却液都与散热器件直接接触。尤其是浸没式液冷系统，其直接将散热器件浸泡在冷却液里。而常规光模块的设计都基于风冷的应用场景，在浸没或喷淋的液冷环境下可能出现失效的情况，因此需要针对液冷环境的应用进行设计并规范相关技术要求，以保障液冷场景下的数据传输。

2) LPO 方案。LPO 通过线性直驱技术替换传统的 DSP，将功能集成到交换芯片中，只留下驱动芯片（Driver）和跨阻放大器（TIA）芯片。LPO 光模块中用到的 Driver 和 TIA 芯片性能也有所提升，从而实现更好的线性度。

3) CPO 方案。CPO 是指把光引擎和交换芯片共同封装在一起的光电共封装，没有采用可插拔光模块的形式。这种方式能够使得电信号在引擎和芯片之间更快地传输，缩短了光引擎和交换芯片间的距离，有效减少尺寸、降低功耗、提高效率。

4) 硅光技术。硅光技术是一种基于硅光子学的低成本、

高速光通信技术，利用基于硅材料的互补金属氧化物半导体（CMOS）微电子工艺实现光子器件的集成制备。该技术结合了 CMOS 技术的超大规模逻辑、超高精度制造的特性和光子技术超高速率、超低功耗（CPO/LPO 架构下）的优势。

5) 薄膜铌酸锂技术。相比于传统体材料铌酸锂，薄膜铌酸锂波导可以制作亚微米尺度的光波导，不但提高了器件的集成度，而且大大提高了对光场的限制能力，增强了光场和铌酸锂材料的相互作用，因而可以实现超高电光带宽、超低驱动电压以及超低光学损耗。该技术有望在电光调制器领域掀起一场革命。

6) 相干下沉。在数据中心本身的范围内，强度调制直接检测（IM-DD）技术仍然占主导地位。认识到 IM-DD 在满足不断发展的需求方面的局限性，业界正在探索“相干精简版”解决方案（一种专门为较短距离的数据中心连接而设计的相干技术方案）。

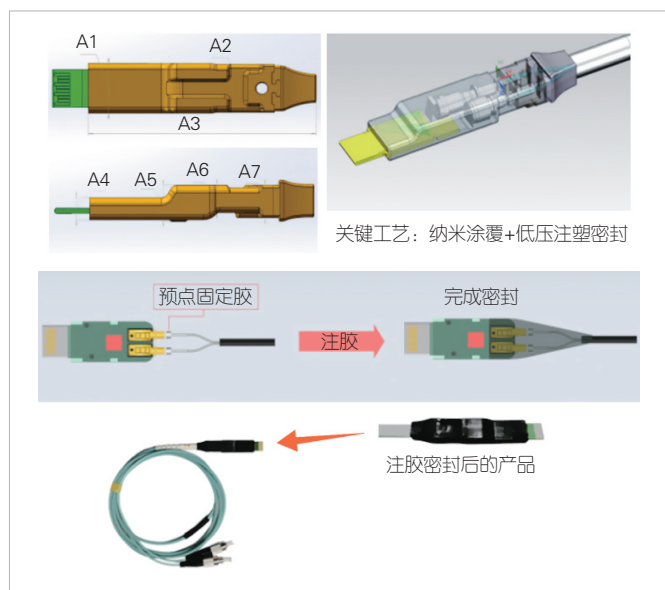
7) 光进光出（OIO）技术。OIO 是一种基于芯片的光互连解决方案，与计算芯片（中央处理器、图形处理器、张量处理器）集成在同一封装中，旨在实现分布式计算系统中它们之间的无缝通信（跨板、机架和计算行），其带宽密度、能源成本和延迟与封装内的电气互连相当。具体产品形态包括板载光学（OBO）、NPO、CPO 及未来可能部署的光电子集成电路（OEIC）。OIO 有望提供百倍以上通道密度，每个通道可以提供百倍以上带宽，很可能成为交换芯片向 100T、200T 乃至更高容量演进的必要技术。

2.1 液冷光模块

当数据中心中采用浸没或者喷淋式液冷系统时，板上芯片封装（COB）工艺的光模块在液体环境中无法直接使用。在液冷数据中心中宜使用蝶形封装 BOX 和同轴封装 TO-CAN 两种气密性封装光组件，并进行模块的整体液密封装，或者对光电部分 COB 封装的光模块进行全塑封设计（液密封装），以有效防止液冷环境中的冷却液对光模块光路部分造成干扰。图 1 展示了易飞扬液冷光模块的密封工艺，该工艺主要通过纳米涂覆和低压注塑密封实现^[2]。

为满足高性能计算和通信应用需求，提高系统性能、能效、可靠性，液冷光模块技术还需要在以下方面实现突破：

- 1) 液冷光模块的技术要求；
- 2) 液冷光模块的封装方式和密封性技术；
- 3) 液冷光模块的液冷测试环境和可靠性测试环境的配置、搭建以及对应测试方法；
- 4) 液冷光模块的运维管理相关技术；
- 5) 液冷光模块的布线连接技术。



▲图1 易飞扬 25G 液冷光模块示意图^[2]

目前，市场上已经有多个厂家发布液冷光模块产品。易飞扬发布了支持浸没式或者喷淋式液冷散热方案的液冷光模块，涵盖 25 Gbit/s、100 Gbit/s 和 200 Gbit/s 速率；海光芯创也在 2023 年的光连接大会现场上演示了自研的 200G QSFP56 SR4 液冷光模块；光迅科技已经可以提供全套的液冷光模块产品，目前已经批量交付 100 Gbit/s 液冷光模块；海信宽带的 25G Pigtail-SFP28 封装液冷光模块于 2023 年已批量生产。另外，华工正源、中天科技也在进行液冷光模块的研究。

2.2 LPO 方案

虽然 DSP 具有数字时钟恢复功能和色散补偿功能，能够以较低的误码率实现信号恢复，但它也带来了较高的功耗和成本开销。例如，在 400G 光模块中用到的 7 nm DSP，其功耗约为 4 W，占到了整个模块功耗的 50% 左右。从成本的角度来看，400G 光模块中，DSP 的物料清单（BOM）成本约占 20%~40%。为了降低功耗和成本，并满足高

速、高密度光通信连接，以及光网络灵活性和可升级性的需求，LPO 应运而生。LPO 方案不采用 DSP/CDR 芯片，而是将相关功能集成到设备侧的交换芯片中。如图 2 所示，该光模块只留下了线性度较高的 Driver 和 TIA，并分别集成连续时间线性均衡（CTLE）和均衡（EQ）功能，用于对高速信号进行一定程度的补偿^[3]。

LPO 具有突出的优势：低功耗、低成本、低延时、易维护。具体表现在：

1) 低功耗。去掉 DSP 后，光模块功耗大幅度下降。相比于可插拔光模块，LPO 的功耗下降约 50%，这与 CPO 的功耗接近。Arista 采用 Linear-Drive，使硅光、垂直腔面发射激光器（VCSEL）、薄膜铌酸锂不同光学方案的功耗均下降 40% 左右。

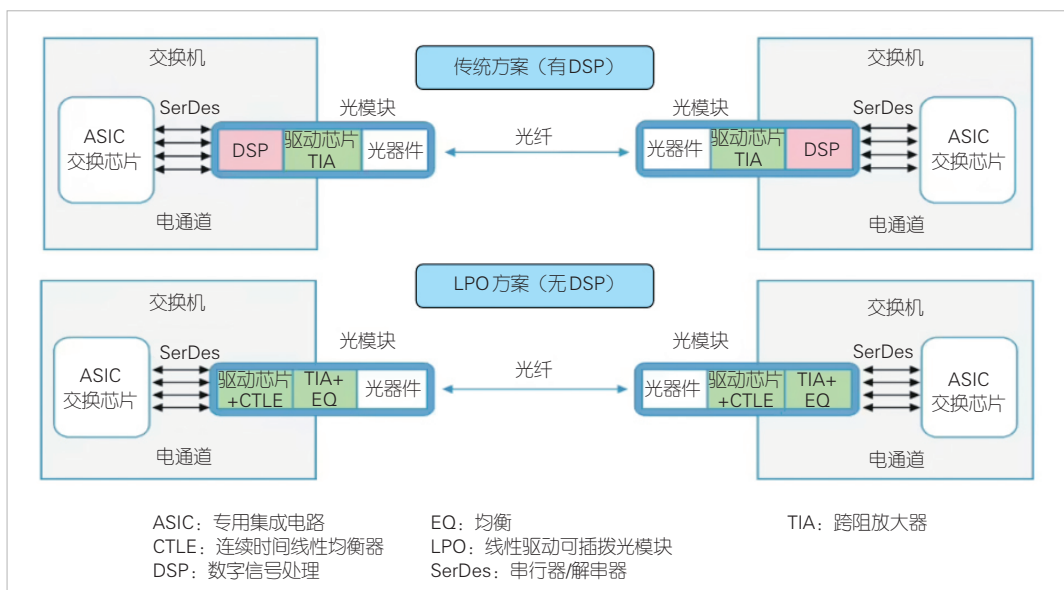
2) 低成本。DSP 的 BOM 成本约占 20%~40%。去除 DSP 后，虽然 Driver 和 TIA 集成了 EQ，局部成本略有增加，但整体成本还是下降的。以 800G 光模块为例，使用 LPO 系统可以使总成本下降大约 8%。

3) 低延时。去除 DSP 会使得光模块减少一个处理过程，数据的传输时延也随之下降。这对人工智能（AI）计算和超级计算场景来说尤为重要。

4) 易维护。在 LPO 方案中，光模块的封装形式没有显著改变，采用可插拔设计，这样方便插入和拔出光学模块，使得光学连接更加灵活便捷。这种设计简化了光纤布线和设备维护，提高了系统的可管理性和可维护性。

然而，当前 LPO 仍存在两大缺点：

1) 通信距离短，应用场景受限制。去除 DSP 后会导致



▲图1 LPO 光模块示意图^[2]

系统的误码率提升,进而导致传输距离变短。

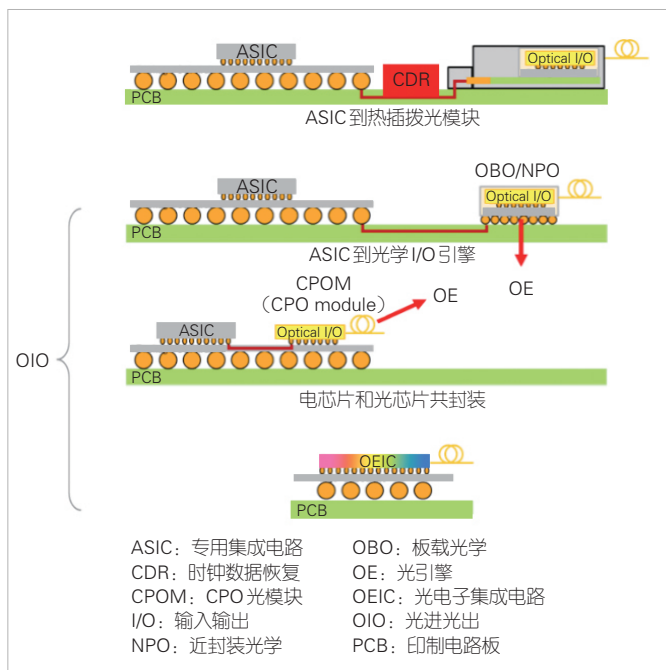
2) 标准化刚起步,互联互通存在挑战。目前由于交换机厂商不倾向于改动交换机,因此光模块需要适应各个通路之间不同的情况。

这里我们给出全球主要企业的LPO部署进度:英伟达披露了在其内部人工智能集群中部署LPO的计划,将在2024年量产LPO光模块;Meta公司预计在2024年上半年导入LPO技术;Arista公司在2023年光纤通信会议和展览会(OFC 2023)上首次展示了有关降低LPO功耗的数据,在网络研讨会上分享了他们在LPO方面的最新研究结果;新易盛、剑桥科技等已发布相关产品;中际旭创已进行相关产品开发;海信宽带已推出800G线性互联光缆。

2.3 CPO方案

CPO是在成本、功耗、集成度各个维度上优化数据中心的光电封装方案。如图3所示,CPO将光模块不断向交换芯片(ASIC芯片)靠近,缩短芯片和模块之间的走线距离,最终将光引擎和电交换芯片封装成一个芯片^[4]。在理想情况下,CPO可以逐步取代传统的可插拔光模块,将硅光子模块和超大规模CMOS芯片以更紧密的形式封装在一起,从而使系统成本、功耗和尺寸都得到进一步优化。如图4所示^[5],按照物理结构分类,CPO可分为3种技术形态:2D平面CPO、2.5D CPO和3D CPO。

1) 2D 封装的CPO技术



▲图3 CPO封装示意图^[4]

基于2D封装的CPO技术是将光子集成电路PIC和集成电路并排放置在基板或PCB上,通过引线或基板布线实现互连。2D封装的优点是易于封装、灵活性高。电子集成电路EIC和光子集成电路PIC都可以使用不同的材料、不同的工艺单独制作。根据芯片和基板互连方式不同,基于2D封装的技术发展出了基于引线键合的CPO、基于倒装的CPO、基于扇出型晶圆级封装技术的CPO 3种技术路径。

2) 基于2.5D封装的CPO技术

2.5D封装将EIC和PIC均倒装在中介层(Interposer)上。通过中介层上的金属互连PIC和EIC,中介层与下方的封装基板或PCB板相连。根据所用转接板的材料不同,基于2.5D封装的技术发展出了基于玻璃转接板的CPO、基于硅转接板的CPO和基于嵌入式多芯片互连桥接3种技术路线。

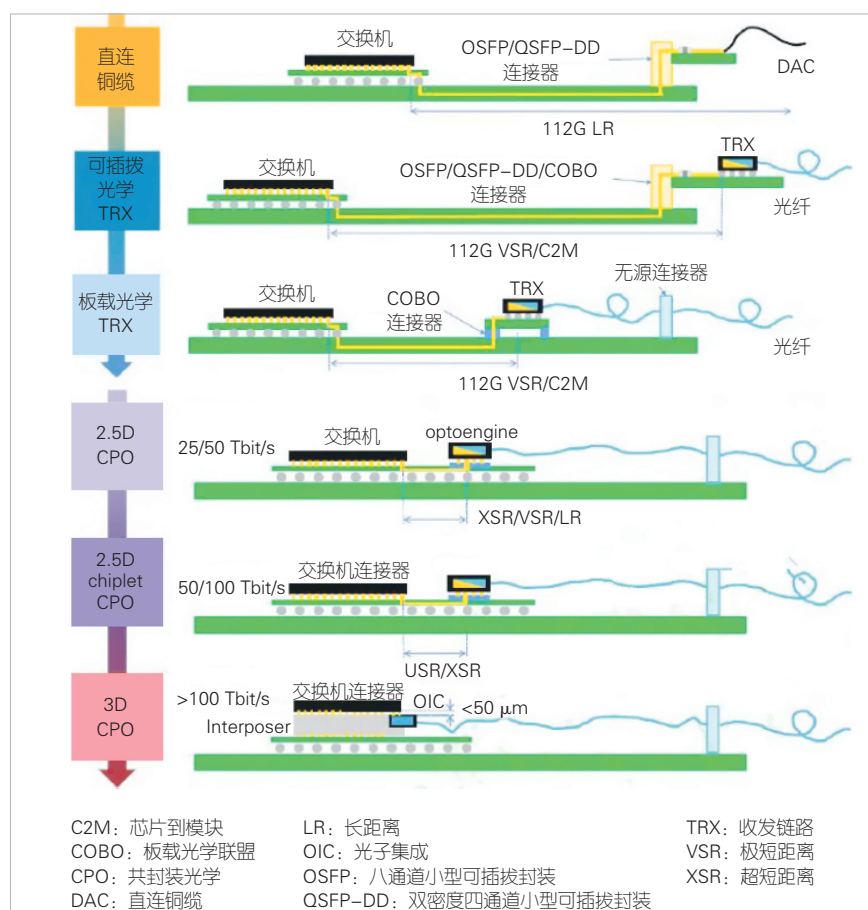
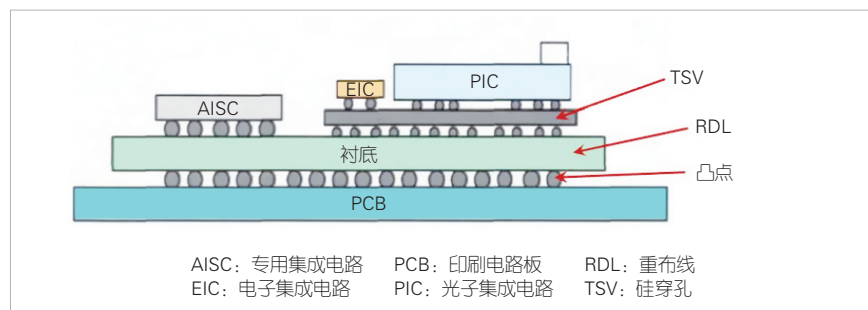
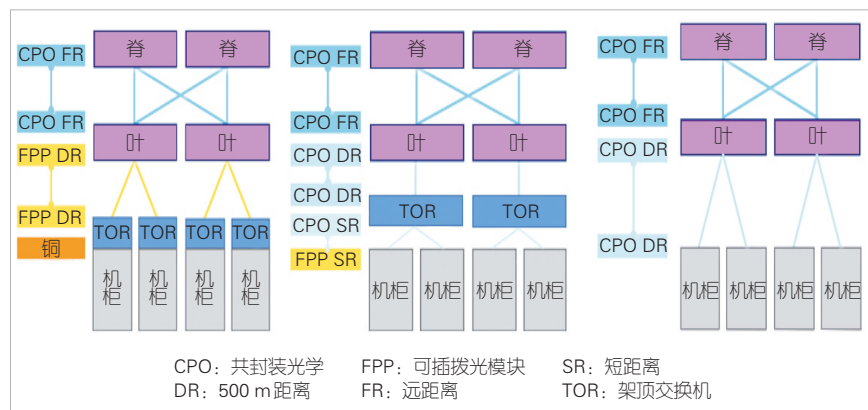
3) 基于3D封装的CPO技术

3D封装技术将光电芯片进行垂直互连,可以不仅能实现更短的互连距离、更高的互连密度和更好的高频性能,还能实现更低的功耗、更高的集成度和更紧凑的封装。基于3D封装的CPO技术是目前CPO技术研究的热点。

CPO技术将增加先进封装工艺需求。如图5所示,目前封装工艺是限制CPO技术发展的主要因素。CPO技术在封装过程中会用到硅穿孔(TSV)、凸点(Bumping)和重布线(RDL)等先进封装技术^[6],对传统光模块封装厂商提出新的挑战。

CPO目前处于产业化初期,在工艺、仿真以及测试等方面仍面临很多挑战。封装工艺能力是制约CPO发展的重要因素,涉及TSV、RDL等多种先进复杂的封装技术。每一种封装技术都有利弊,因此我们需要不断探寻最可靠的方案。散热问题对于CPO来说是一个重大挑战。CPO中放置光器件和电器件的空间十分有限,并且光器件对热特别敏感。CPO标准工作组做的模拟仿真表明,在风速5 m/s的条件下,当采用16个的CPO模块和1个开关芯片模型设计时,开关芯片的温度达到151.76℃,几乎无法正常工作。主流的热可插拔模块维修简单,但作为不可插拔的CPO封装技术维修难度较高,因此测试、良率以及可靠性问题成为CPO产业化的关键。由于光芯片是直接和电芯片通过先进封装工艺封装在一起的,如果某颗芯片发生了损坏,整个模块就无法正常工作。这不仅使良率下降,也给测试带来了诸多困难。CPO光学器件的通道密度与以往任何子组件不同,这意味着现有的测试解决方案需要改进。

CPO的发展需要产业链协同推进,这将考验光模块/光引擎厂商的长期综合实力。如图6所示,CPO的技术路线优化本质上是对整个网络架构的优化^[5],需要数据中心整体产

▲图4 CPO技术路线^[5]▲图5 基于带TSV的PIC转接板的共封装光学(CPO)^[6]▲图6 CPO与数据中心网络架构优化路径^[5]

产业链的协同推进。其中,在现有光模块产业链的基础上,有些环节还需要交换芯片及设备厂商、各元器件厂商共同参与。因此CPO的发展本质上是对光模块/光引擎厂商综合实力的长期考验。

2020年以来,CPO逐渐从学术研究成果转变为市场需求产品。如表1所示,博通、Cisco、Marvell等行业内龙头企业均已推出多款基于CPO样品,其他企业也在积极地布局相关产品,并推进CPO技术标准化。云服务厂商Facebook和Microsoft创建了CPO联盟,旨在打造一个平台,以吸引各细分行业龙头企业加入联盟,推动CPO标准的建立和产品的发展。中国企业则普遍较晚进入CPO领域,在产品开发进度及技术研究方面存在明显的差距。中国计算机互连技术联盟(CCITA)牵头的CPO标准是当前中国唯一原生的CPO技术标准。该标准旨在结合目前全球光互连技术发展,联合中国光模块、光收发芯片、电驱动放大芯片、光源、连接器等厂商,共同打造更加适合中国的CPO标准。光迅科技、中际旭创、华工科技等都已开始涉足光电共封领域,但由于起步较晚,目前还没有CPO相关的产品推向市场。

2.4 硅光技术

相对于传统光模块,硅光模块在高速率领域具有高集成度、低成本、低功耗的显著优势。硅是用量最大的半导体晶圆材料,具有低成本和加工工艺成熟的优势。硅光基于硅和硅基衬底材料,通过CMOS工艺进行光器件开发和集成。硅光模块产生的核心理念是以光代电,即利用激光束代替电子信号进行数据传输。传统光模块采用分立式结构,制造过程中需要依次封装电芯片、光芯片、透镜、对准组件、光纤端面等器件,部件物料多。如图7所示,硅光模块将激光器、调制器、探测器等光/电芯片都集成在硅光芯片上。传统器件中的透镜和大型组件都被取代,陶瓷、铜等材料用量大幅降低,晶圆、硅光芯片等电

▼表1 全球主流企业的CPO技术研究进展

公司	进展情况
博通	2023年10月联合 Micas Networks(云母网络)在美国 OCP 全球峰会上展示了基于博通的 Bailly CPO 首款 51.2T (CPO)交换机。该产品的每个光引擎为 3.2T(2×8 PCS),集成 CMOS 电芯片和 MUX/DMUX,整体相对领先
Cisco	2023年3月在 OFC 2023 上展示基于 Silicon One G100 ASIC(25.6T)CPO 交换样机。51.2T+CPO 的 DEMO 样机还在开发中,预计会在 OFC 2024 上展示
英伟达	积极推进芯片光互连策略,通过战略投资 Ayar Labs、TSMC 构建硅光平台、芯片共封装能力,计划 2025 年左右通过 CPO 技术实现 GPU 芯片与 NVSwitch 芯片之间的光连接
Marvell	2022 年推出基于 2.5D CPO 技术的 12.8 Tbit/s Teralynx7 交换机,2023 年推出由超低延迟 Marvell Teralynx10 51.2 Tbit/s 交换芯片和业界首款 PAM4 1.6 Tbit/s 光电平台 Marvell Nova 组成的新平台
光迅科技	在 OFC 2023 上发布可以支持 3.2T CPO 光引擎的自研光源模块
中际旭创	2021 年开始进行关键技术的预研,持续打造先进光子芯片产业化技术平台和 2.5D、3D 混合封装平台
锐捷网络	2021 年发布首款基于 112G SerDes 交换芯片和由 16 个 1.6 Tbit/s CPO 模块组成的 25.6 Tbit/s CPO 交换机;2022 年将 CPO 模块从 1.6 Tbit/s 升级到 3.2 Tbit/s,并发布了 51.2 Tbit/s 的 CPO 交换机

CMOS: 互补金属氧化物半导体

CPO: 共封装光学

DMUX: 解复用器

DR: 500 m 距离

ER: 加长距离

FR: 远距离

GPU: 图形处理器

LR: 长距离

MUX: 复用器

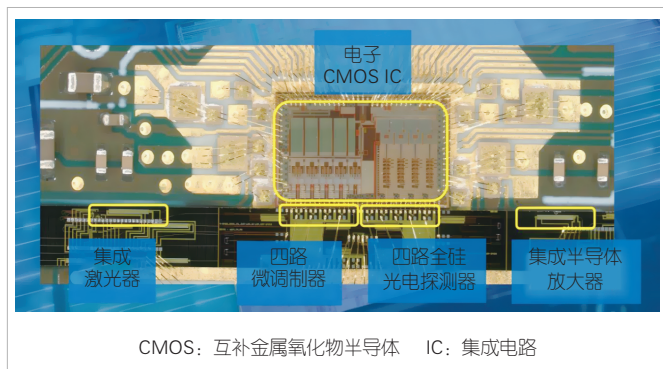
OFC: 光纤通信会议和展览会

PAM4: 四电平脉冲幅度调制

SerDes: 串行器/解串器

SR: 短距离

ZR: 最长距离



▲图7 Intel硅光模块样品示意图

子材料占比上升。光模块价值向硅光芯片、硅光引擎转移。传统光模块制造过程中封装工序较为复杂,需要投入较多人工成本,而硅光芯片高度集成,组件与人工成本也相对降低,对下游封装厂或制造商的要求也会降低。

硅光技术的发展可以分为4个阶段。第1阶段,硅基器件逐步取代分立元器件,即将硅作为光通信底层器件,并达到工艺的标准化水平。第2阶段,集成技术从耦合集成向单片集成演进,实现部分集成,再通过不同器件的组合,把这些器件集成不同的芯片。第3阶段,光电一体技术融合,实现光电全集成化。把光和电都集成起来,可实现更加复杂的功能。第4阶段,硅光技术实现可编程芯片。这一阶段器件能够分解为多个硅单元排列组合,局针化表征类。该种通过编程来改变内部结构的芯片,可自定义全功能。目前硅光技术已经发展到了第2个阶段。硅光子核心器件主要是以硅半导体材料的光有源及无源器件:硅基激光器(负责将电信号转化成光信号)、硅基光调制器(负责将光信号带宽提升)、

硅基光探测器(负责将光信号转化成电信号)、光波导(负责光信号在硅基材料上传输)、光栅耦合器(负责与对外连接的光纤对准降低插损)等。

1) 硅基激光器。硅基激光器指集成在以硅为衬底的光芯片上的激光器。常用的硅基激光器按照结构可分为 VCSEL 芯片、分布式布拉格反射激光器(DBR)芯片、分布式反馈激光器(DFB)芯片和电吸收调制激光器(EML)芯片。制作材料主要以 III-V 族半导体材料为主。对于 III-V 族激光器与硅光芯片的耦合,主流设计方案主要有片上倒装焊集成、异质键合集成和直接外延生长集成。异质键合集成和直接外延生长集成是未来实现硅光大规模生产的可行方案。

2) 硅光调制器。硅光调制器指集成在硅光芯片上的调制器。硅光调制器集成度高、消光比较高、损耗低、驱动电压小,但线性度差,因此目前业界多使用混合集成调制器。通过异质键合、外延等技术,将成熟的铌酸锂调制器、InP 调制器等集成到硅基上,可实现微米级大小,调制效率为全硅调制器的 5 倍以上。

3) 硅基光探测器。光电探测器将接收的光信号转变为电信号。由于硅对 1.1 μm 以上的光波透明,单体硅无法制作探测器。目前集成在硅基片上的高频探测器主要有混合集成 III-V 族和硅锗混合探测器。前者耦合效率高、灵敏度高、响应快;后者性能优越,器件制备技术与 CMOS 工艺兼容,更适合大规模集成,是目前的主流方案。

就硅光技术在通信设备中的具体应用而言,目前除了激光器外,光模块中大部分器件的制造都已实现,典型的如光波导、外调制器件、雪崩光电二极管(APD)接收器等。不

过由于硅光技术产业尚处于起步阶段，业内并未形成权威的
行业规范与技术标准，各主流厂商采用硅光技术设计生产元
器件时采用的技术路线不尽相同，因此最终的技术方案还有待
优化完善。

如图8所示，从工艺角度来看，硅光可以分成单片集成
和混合集成。目前混合集成使用较广，但是单片集成性能更
优，是未来发展趋势。单片集成是指将光子学组件直接集成
到同一块硅芯片上，包括光源、光调制器、波导、耦合器等
光学元件，从而形成一个紧凑的光学电路。单片集成方式的
优势在于可以减小尺寸，提高集成度，降低制造成本。混合
集成是指将硅芯片与其他材料的光学组件结合在一起，即将
电子器件（硅锗、CMOS、射频等）、光子器件（激光/探测
器、光开关、调制解调器等）、光波导回路集成在一个硅芯
片上。其中，硅芯片主要负责电子部分的处理，而其他材料
的光学元件则负责光的生成和调制。混合集成的优势在于可
以利用硅芯片的电子器件和其他材料的优异光学特性，实现
更高效的光通信和传感应用。目前来看，光器件如波分复用
器、变换调谐器等已经可以实现单芯片集成，而光模块尚需
要混合集成。

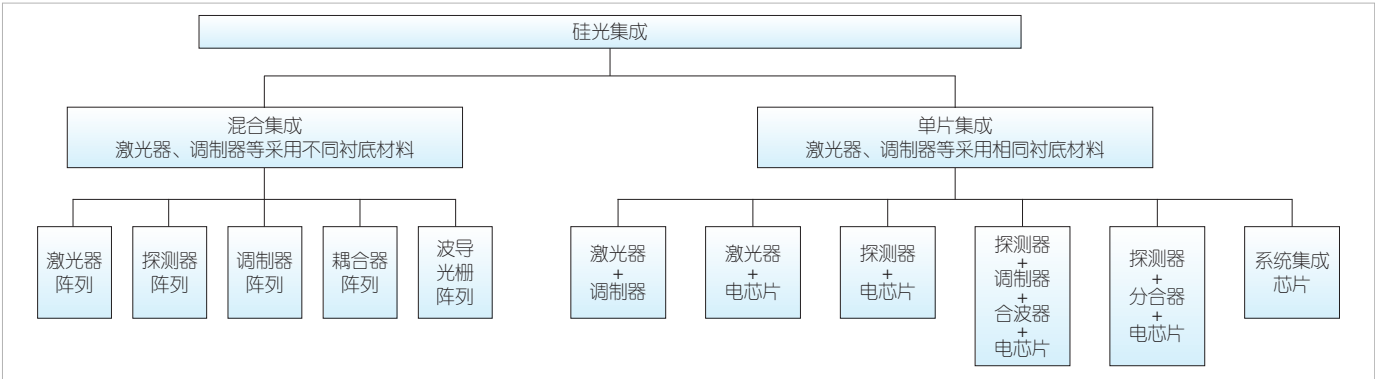
虽然硅光集成技术具有巨大的市场前景，但目前仍然面
临诸多挑战：

1) 硅光器件性能问题。目前的硅光技术已可以替代很
多传统的光器件，但是还有一些需要克服的技术难题，比
如：如何减少硅波导的损耗，如何实现波导与光纤的有效耦
合，如何克服温度对于功率和波长稳定性的影响等。这些技
术难题会影响到硅光技术的普及以及在数据中心场景中的
应用。

2) 测试流程复杂。与常规的大规模集成电路芯片不同，
光电芯片本身成本高、制造流程多、工艺复杂、废品率高，
因此需要先在晶圆上进行测试和筛选，然后再和其他电芯片
进行集成，以避免后期由残次芯片造成的不必要的封装
成本。

3) 标准化方案缺乏。硅光芯片在各个环节都缺少标准
化方案，例如：设计环节需要使用专用的EDA工具，制造
与封装环节缺乏提供硅光工艺晶圆代工服务的厂家。这使
得硅光技术大规模产业化变得更加困难。

硅光技术主要布局公司如表2所示。当前，台积电计划
携手博通、英伟达等共同开发硅光子技术、共同封装光学元
件等新产品。制程技术从45 nm延伸到7 nm。台积电已建
一支由约200名专家组成的研发团队，专注于利用硅光子技
术开发未来芯片。



▲图8 硅光集成方案

▼表2 主流公司布局硅光技术对比

公司	硅光技术进展
Marvell	2022年推出业界首款800 Gbit/s 多模平台解决方案,同年量产用于数据中心的400G DR4硅光子平台
中际旭创	400G硅光模块已进入市场导入阶段,正在接受海外客户认证;800G硅光模块已开发成功,并向海外客户送样;400G/800G的硅光模块已采用自研的硅光芯片
熹联光芯	100G硅光模块已实现规模化量产,400G 光学引擎及硅光模块正处于多个客户认证测试中;在张家港经开区建设中国第一条硅光芯片及封测生产线,项目总投资达20亿元
华工科技	2021年400G硅光芯片已实现量产,2022年发布800G硅光芯片并实现小批量生产
光迅科技	2018年发布100G硅光收发芯片并正式投产使用,200G/400G硅光数通模块已经开始出货,800G产品已开始向客户送样
华为	2012和2013年分别收购英国光子集成公司CIP和比利时硅光子公司Caliopa,2019年后累计投资10余家光芯片产业链相关企业,比如熹联光芯、微源光子、长光华芯、芯视界等

2.5 薄膜铌酸锂技术

电光调制器是超高速数据中心和相干光传输的核心光器件。通过调制将通信设备中的高速电子信号转化为光信号，电光调制器成为光通信系统中不可或缺的一环。如表3所示，目前光调制技术主要基于硅光、磷化铟和铌酸锂3种材料的电光调制器。其中，铌酸锂电光系数显著高于磷化铟，而硅没有直接电光系数，因而铌酸锂调制器是高带宽光电信息处理系统中的关键器件。传统铌酸锂电光调制器为体材料铌酸锂调制器。体材料铌酸锂调制器具有带宽高、稳定性好、信噪比高、传输损耗小、工艺成熟等优点，但在传输速率需求不断提升的形势下，体材料铌酸锂调制器也在一些性能上遭存在瓶颈，而且体积较大，不利于集成。新一代薄膜铌酸锂调制器芯片技术将解决这些问题。铌酸锂材料具有“光学硅”之称。通过最新微纳工艺制备出的薄膜铌酸锂调制器，具有高性能、小尺寸、可批量化生产且与CMOS工艺兼容等优点。

与磷化铟相比，薄膜铌酸锂在长距离/超长距离的产品封装上具有优势。它可以采用非气密封装，从而降低成本，而磷化铟则需要气密封装。在性能方面，薄膜铌酸锂对比硅光和磷化铟都有明显优势。铌酸锂薄膜调制器利用容性电极和石英衬底，同时实现低微波损耗和光电同步传输。容性行波电极不仅可以降低金属微波损耗，还可以降低微波速率，使光波和微波在调制区域以相同的速率传输，光和电达到完

美匹配。如图9所示，当前800G主流方案包括DR8/SR8/2FR4等方案。无论是单模还是多模方案，都使用单通道100 Gbit/s速率。向单波200/400 Gbit/s速率演进时，薄膜铌酸锂的大带宽优势将更加突出。未来对于成本、功耗、性能等要求会越来越高。薄膜铌酸锂调制器凭借功耗、成本、性能等方面的优势，有望迎来快速发展。

薄膜铌酸锂目前主要面临两个方面的挑战：1) 低成本制备工艺。薄膜铌酸锂目前成本较高，而商用后的市场需求与单位成本息息相关。因此，如何解决产业化后的成本问题是关键。2) 铌酸锂的晶圆尺寸。目前铌酸锂的晶圆尺寸以4英寸（1英寸=2.54 cm）和6英寸为主。8英寸和12英寸能否实现产业化与后期的成本相关。

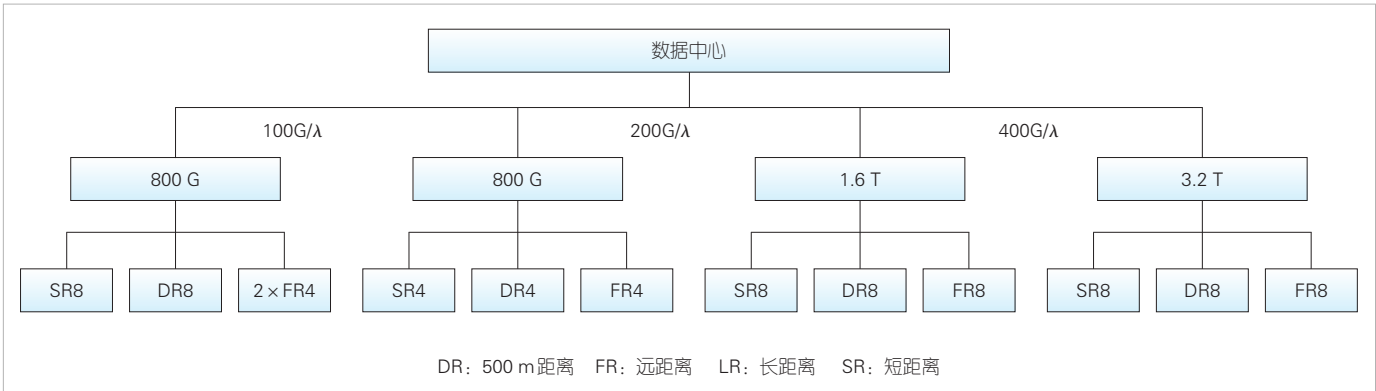
考虑到薄膜铌酸锂材料优异的性能，如表4所示，当前主要光模块厂商均已布局铌酸锂技术。

2.6 相干下沉

相干技术是长距光传输采用的技术。由于直检技术面临色散、四波混频等挑战，传输距离不断缩小。业界出现了相干技术下沉到数据中心互联的发展趋势。相干技术传输性能好，可以灵活地采用光数字信号处理（oDSP）技术进行色散补偿，但是成本和功耗较高。为了降低成本和功耗，许多高校及企业提出了Coherent-Lite的概念，并提出O波段相干、灰色（固定波长）激光器和自相干。

▼表3 电光调制器3种材料方案对比

材料方案	主要应用场景	优势	劣势
硅光方案	数据中心内部互联光模块	尺寸小,易集成,成本低,能耗低	调制性能一般,带宽低,插入损耗大
磷化铟方案	中距和长距光通信网络收发模块	尺寸小,集成度高	成本高,良率低,功耗高,偏振器件无法集成
铌酸锂方案	100 Gbit/s以上的长距骨干网相干通信,单波100/200 Gbit/s的超高速数据中心	电光性能好,带宽高,稳定性好,信噪比高,传输损耗小,工艺成熟	传统型尺寸大,不利于集成



▲图9 薄膜铌酸锂应用场景

▼表4 主流公司布局铌酸锂技术对比

公司	铌酸锂技术进展
新易盛	在OFC大会上展示最新800G光模块,为800G光模块功效提供了新的标准参考;DSP直接驱动薄膜铌酸锂,实现了出色的光发射性能
中际旭创	将根据市场需求开发相关产品
联特科技	比较看好硅光和薄膜铌酸锂两项技术,会并行开发这两项技术;公司年报已披露基于薄膜铌酸锂调制技术的800G光模块产品
光迅科技	正在持续研发中,目前已有小批量的供货
铌奥光电	开发的薄膜铌酸锂调制器芯片已实现超低驱动电压($V_{\pi} < 1\text{ V}$)和超大带宽(最新记录为260 Gbaud),产品功耗较硅光调制器等显著降低

DSP: 数字信号处理 OFC: 光纤通信会议和展览会

1) O波段相干。目前数据中心内部光信号传输的波长都是O波段,而传统的相干光通信是C波段。C波段最大的好处是光纤功率损耗最小,所以非常适合长距离的光纤传输;但同时C波段的光纤色散比较大,需要用DSP进行色散和各种损伤的补偿。而在O波段,虽然光纤功率损耗变大了,但好在传输距离并不远,功率衰减不多,所以也不需要光放大器。同时在O波段光纤的色散最小,这样的话就不需要DSP来补偿色散。因此,DSP的功能可以进一步简化,这使得系统成本和功耗进一步降低。

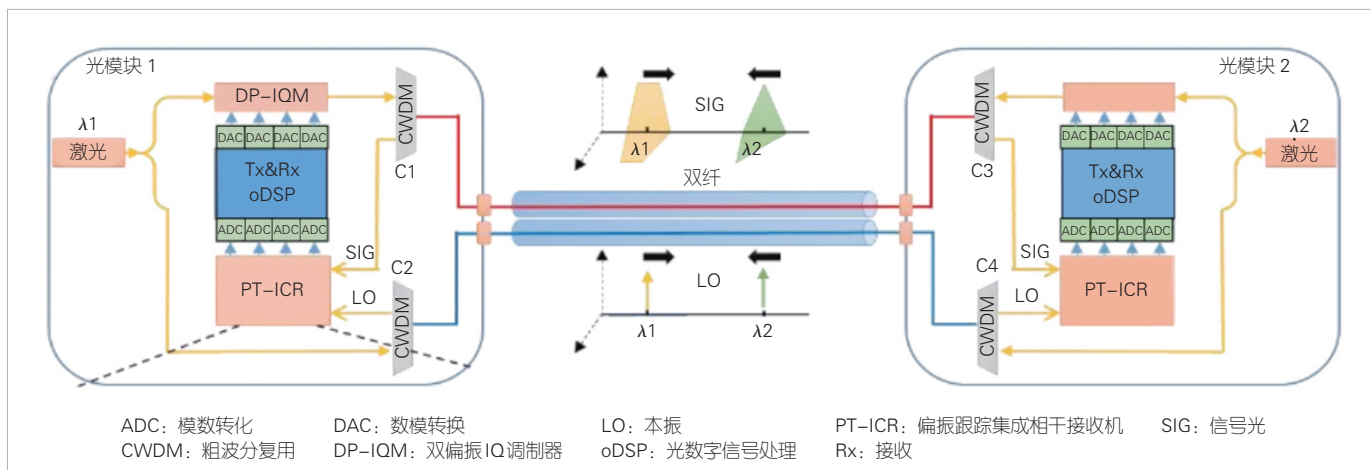
2) 灰色(固定波长)激光器。由于数据中心内部架构不需要在光纤中进行密集的波长传输,因此可以使用灰色(固定波长)激光器。它的使用消除了对波长调谐以及相关控制电路和算法的需求,简化了实施过程,降低了操作的复杂性。

3) 自相干。传统的相干方案是异源相干,本振(LO)光要求与信号载波频率相位对齐,而自相干是同源相干,即一根光纤双向传输数据,一根光纤双向传输LO,如图10所示^[7]。这个架构的好处是:不需要使用昂贵的激光器,无温控热电制冷器(TEC),无波锁,可有效降低模块成本和功耗。此外,接收机信号不存在频偏和相噪问题,DSP可以得到进一步简化。

随着传输速率提升至800G,相干技术方案在80 km传输距离的基础上将进一步向10 km等更短距离拓展应用。当传输速率达到1.6T时,相干会进一步下沉到2 km。IEEE 802.3正在标准化800G/1.6T以太网接口,包括单通道100G和200G两路不同传输距离的接口。值得一提的是,2023年,IEEE 802.3dj就800G 10 km应用是采用强度调制和直接检测(IMDD)还是采用相干技术进行了激烈讨论。最终,IEEE 802.3dj决定为800G 10 km设定两个项目目标,分别采用不同的技术解决方案。可以看到,随着单通道速率的提升,相干技术正在不断下沉,相关应用场景也在不断拓展。

3 结束语

液冷技术具备超高能效、超高热密度等优点,是解决数据中心散热压力和节能挑战的重要途径。液冷光模块是大势所趋。LPO具有短距离、低功耗、低时延等特性,能够适配AI计算中心。由于可以直接应用于目前成熟的光模块供应链,LPO在高线性度TIA/Driver厂商的大力推动下或可快速落地。CPO方案通过交换机光电共封装降低成本和功耗,是未来实现高集成度、低功耗、低成本、超高速率模块应用方面是综合最优的封装方案。虽然硅光产品的研发投资和销售额仍小于III-V族材料,在产品性能、工艺、成本等方面仍

▲图10 自相干示意图^[7]

面临一定挑战,但基于其在成本和功耗方面的优势,硅光技术有望成为未来光器件的主流技术。薄膜铌酸锂是超高速数据中心和相干光传输的核心光器件,具有高性能、低成本、小尺寸、可批量生产且与CMOS工艺兼容等优点,是未来高速光互连极具竞争力的解决方案。当数据中心内部连接的要求超过1.6T并走向3.2T时,色散损伤和链路预算将对直接检测(IM-DD)解决方案提出更高要求。因此,相干精简版解决方案有望成为大批量3.2T数据中心内互联应用的有力竞争者。

致谢

感谢中兴光电子技术有限公司沈百林、熊孝海等专家对本研究的帮助!

参考文献

- [1] Cisco. 2020 Cisco and/or its affiliates [EB/OL]. [2024-01-15]. https://www.cisco.com/c/dam/global/en_sg/partners/assets/networking-academy-cybersecurity/Josh-mccloud-becybersmart-presentation.pdf
- [2] 中国通信标准化协会. 用于液冷系统中的光模块研究报告 [R]. 2021
- [3] 西南证券人工智能专题研究: LPO 光模块 [EB/OL]. [2024-01-15]. https://pdf.dfcfw.com/pdf/H3_AP202312061613416719_1.pdf?1701854056000.pdf
- [4] IPEC. IPEC 成功立项 OIO 研究项目, 探索下一代数据中心交换芯片技术演进 [EB/OL]. [2024-01-15]. <https://www.ipec-std.org/zh/5219.html>
- [5] MINKENBERG C, KRISHNASWAMY R, ZILKIE A, et al. Co-packaged datacenter optics: opportunities and challenges [J]. IET optoelectronics, 2021, 15(2): 77-91. DOI: 10.1049/ote2.12020
- [6] 卞玲艳, 曾艳萍, 蔡莹, 等 大数据时代光电共封技术的机遇与挑战 [J]. 激光与光电子学进展, 2024, 61(11). DOI: 10.3788/LOP231348
- [7] GUI T, WANG X F, TANG M, et al. Real-time demonstration of

600 Gbit/s DP-64QAM selfomodyne coherent bi-direction transmission with uncooled DFB laser [C]//2020 Optical Fiber Communications Conference and Exhibition (OFC), 2020, IEEE. DOI: 10.1364/OFC.2020.Th4C.3

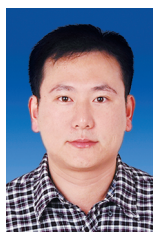
作者简介



张平化, 中兴光电子技术有限公司工程师; 主要从事光模块规划工作。



王会涛, 中兴光电子技术有限公司规划总工; 长期从事光通信系统、光电子器件产品和技术的研究及规划工作; 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。



付志明, 中兴光电子技术有限公司工程师; 主要从事光模块规划工作。

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

第 30 卷 总第 174 期

2024 年 2 月 第 1 期

主管：安徽出版集团有限责任公司

主办：时代出版传媒股份有限公司

深圳航天广宇工业有限公司

出版：安徽科学技术出版社

编辑、发行：中兴通讯技术杂志社

总编辑：王喜瑜

主编：王利

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

《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

电子信箱：magazine@zte.com.cn

电话：(0551) 65850139

发行方式：自办发行

印刷：合肥添彩包装有限公司

出版日期：2024 年 2 月 29 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