



中移智库



中国移动 研究院
China Mobile CMRI



6G 智能空口技术白皮书

(2026 年)

发布单位：中移智库

编制单位：中国移动通信有限公司研究院、北京航空航天大学、北京理工大学、清华大学、中国科学技术大学、北京三星通信技术研究有限公司、大唐移动通信设备有限公司、中关村泛联移动通信技术创新应用研究院、中山大学、上海大学

前言

我们正站在通信技术代际跃迁的关键节点。从 1G 到 5G，移动通信以十年一代的节奏，推动了从个人连接到万物互联的深刻变革。然而，面向 2030 年及未来的“万物智联”时代，全息通信、数字孪生、自动驾驶、智能体等新兴应用对网络提出了更高要求：极致的频谱与能效、内生智能、通感一体、以及对复杂环境的自主认知能力。6G 作为未来智能社会的关键基础设施，其核心引擎——无线空口，亟待一场根本性变革。

回首过往，空口技术的演进之路辉煌而艰辛。大规模多输入多输出（Multiple Input Multiple Output, MIMO）、波形多址、编码调制等关键技术，已将传统通信理论的潜力挖掘至深。我们不得不面对一个日益严峻的现实，那就是**依靠规模扩张与线性优化的传统技术路径已渐趋瓶颈**：链路级编码调制的性能已逼近香农极限，更大规模的天线阵列意味着难以承受的导频开销与计算复杂度，密集化、超大规模的网络部署使得系统干扰环境空前复杂，多小区联合处理对回传容量和实时信道获取提出了近乎苛刻的要求。这一切都表明，无线空口技术已步入一个“高原区”，若仅沿着既有范式修修补补，将难以满足 6G 的宏伟愿景。空口技术正呼唤一场范式革命。

人工智能（Artificial Intelligence, AI）技术的兴起，为突破 6G 空口设计瓶颈提供了全新范式。其核心价值体现在三大能力突破上：**一是卓越的复杂系统建模能力**，能从高维数据中挖掘信道特性、器件非线性等深层特征，构建更精确的实用模型；**二是高效的高维空间寻优能力**，运用强化学习等方法，对资源分配、干扰协调等挑战问题给出近最优解；**三是革命性的端到端联合设计能力**，通过统一的神经网络打破模块壁垒，将级联优化转化为端到端学习，实现全局性能增益。这些能力共同驱动空口设计实现从传统范式向智能范式的根本性转变。

然而，将 AI 引入高要求的空口设计面临核心矛盾：性能增益必然伴随算力、数据资源和存储开销等新成本，缺乏科学权衡将导致“AI 吞噬增益”的困境。为了构建绿色高效的 6G 智能空口，并全面提升系统容量、谱效与能效等关键性能，需重点突破以下研究方向：确立空口 AI 设计准则、重塑 AI 与通信双向赋能的接入网架构、构建低开销数据集，以及设计高效模型与空口传输算法。这些课题已成为推动 6G 空口演进的核心研究任务。

在设计准则方面，本白皮书提出“能力-质量-代价”三角均衡设计评估模型，系统权衡三大维度：能力（谱率、能效等通信指标），质量（推理精度、时延、泛化性等模型性能），代价（算力、数据采集、训练推理等相关成本）。这一模型要求在三维约束空间寻求最优解，成为评判空口 AI 技术价值的基本准则。

在顶层架构方面，提出“智能内生”的 6G 接入网架构，旨在从根本上解决 5G 通过外挂式、烟囱式引入 AI 所带来的效率瓶颈与能力割裂。通过新增面向数据全生命周期管理的“数据面”和面向高性能处理的“计算面”，并增强控制面与用户面，在架构层面实现连接、计算、数据与 AI 模型四要素的深度融合与统一灵活编排，为空口 AI 能力的按需部署与高效发挥提供坚实的土壤。

在关键技术方面，本白皮书进行了一系列从理论到实践的创新探索。从低开销空口 AI 数据集的构建方法，到面向 MIMO 信道信息获取、器件非线性补偿、单小区与多小区用户调度与 MIMO 预编码、信源信道联合设计乃至无循环前缀波形等核心环节的创新 AI 模型设计，均提供了解决方案与性能验证。其中很多模型设计是对“能力-质量-代价”均衡理念的深入诠释与工程实践，以更低的复杂度实现更高的系统容量、谱效和能效等指标。

在产业推动方面，坚持技术可行性与标准化并重。通过集中式与分布式大规模 MIMO 样机、超高阶调制样机等一系列原型验证，证明了多项空口 AI 技术在真实环境下的卓越成效。同时，团队深度参与第三代合作伙伴计划（3rd Generation Partnership Project, 3GPP）等国际标准化组织，积极将研究成果融入全球技术规范，旨在为构建全球统一、高效协同的 6G 智能空口标准贡献中国智慧与力量。

本白皮书是团队在 6G 空口 AI 领域探索的成果总结。我们期待与全球产学研界同仁分享见解、凝聚共识、协同创新。我们坚信，以“能力-质量-代价均衡”为准则，以“智能内生”为架构基石，6G 空口将突破传统通信极限，从被动传输管道演进为“认知共生”的自优化与自演进智能系统。这不仅是技术的跃迁，更是从“连接万物”迈向“万物智联”新纪元的关键一步。

未来已来，让我们携手绘制 6G 智能空口的发展蓝图。

本白皮书由中移智库发布，未经书面授权，任何单位或个人不得复制或拷贝本文件之部分或全部内容。

目 录

1. 白皮书背景.....	1
2. 白皮书章节安排和主要技术突破.....	3
3. 6G 空口 AI 技术设计理念.....	5
3.1. 设计思路.....	5
3.2. 优化策略.....	7
4. 6G 智能内生接入网架构设计.....	12
4.1. 6G 智能内生网络架构需求和驱动力.....	12
4.2. 6G 智能内生网络架构总体设计原则.....	13
4.3. 6G 智能内生网络架构设计方法.....	13
4.4. 6G 智能内生总体网络架构设计框架.....	15
4.5. 6G 智能内生接入网架构方案设计.....	17
4.6. CIS-RAN：智能内生接入网架构的演进.....	20
5. 低开销空口 AI 数据集构建.....	21
5.1. CSI 压缩反馈低开销采集方法.....	21
5.2. 低开销 AI 室内定位数据采集.....	24
5.3. AI 接收机低开销数据采集.....	26
6. 空口 AI 传输技术设计.....	27
6.1. 信道信息获取方法.....	28
6.2. 用户调度与 MIMO 预编码.....	44
6.3. 空口 AI 编译码和波形设计方法.....	54
6.4. 空口 AI 多模块联合设计方法.....	58
7. 空口 AI 多模块集成与系统级仿真.....	62
7.1. 业界面临的挑战.....	62
7.2. 思路和方法.....	63
7.3. 成效和先进性.....	64
8. 空口 AI 样机.....	66
8.1. 集中式大规模 MIMO 样机.....	66
8.2. 分布式大规模 MIMO 样机.....	68
8.3. 超高阶调制样机.....	69
9. 总结与展望.....	70
10. 白皮书贡献单位及人员.....	71
缩略语列表.....	72
参考文献.....	74

1. 白皮书背景

为应对 6G 在容量、峰值速率、用户体验速率、谱效和能效等方面的严峻挑战，业界普遍考虑在 5G 基础上通过增加基站密度、天线数、射频通道数、数据流数及系统带宽等方式提升性能。然而，这种资源堆叠模式将显著增加系统复杂度、功耗与成本，难以高效、绿色地满足 6G 多样化需求。

在经典信息论框架下，传统信号处理算法在线性调制解调、编译码及 MIMO 预编码等理想信道条件下表现优异，部分调制技术甚至可逼近香农极限。然而，在实际场景中，传统方法在以下三类情况下面临技术瓶颈，难以提升系统性能：一是难以对复杂通信问题建模，如复杂信道中的噪声与干扰建模、多模块联合优化建模；二是可建模但无法高效求解，如多小区用户调度与分布式 MIMO 预编码；三是可建模并存在算法，但算法受限于硬件实现，如高维矩阵求逆的 MIMO 预编码与检测。AI 技术通过神经网络对复杂问题进行建模与低复杂度求解，有望显著提升系统容量、谱效与能效，成为 6G 空口设计的重要手段。

目前，AI 技术已在编码、调制、多址接入、波形设计、MIMO 预编码、信道估计、信号检测、语义提取与资源调度等多个空口模块中展现出潜力。3GPP 也已启动物理层 AI 标准化工作，为空口 AI 的落地奠定了基础。然而，空口 AI 的研究依然面临如下多重严峻挑战：

1) **缺少科学的设计评估准则**：当前 6G 空口 AI 研究主要聚焦于设计高精度 AI 模型以提升通信能力，普遍忽视了工程实践中所需的算力、复杂度和空口资源等 AI 代价，对模型泛化能力和推理时延等关键 AI 质量指标也缺乏系统性考量。这种过度依赖算力资源追求通信能力提升的研究范式，难以支撑智能化网络的可持续发展。

2) **智能接入网架构缺乏系统设计思路**：现有的 5G 网络架构在数据全生命周期管理和计算服务性能方面均存在不足。面向 6G，如何设计连接、数据、计算多要素协同的接入网架构以实现 AI 与网络双向赋能是一个非常挑战的开放问题。

3) **数据集构建面临挑战**：仿真数据集的特征难以与实际的应用场景匹配，而基于实际测量的数据集构建会带来大量的测量开销。如何以低测量开销和低反馈开销获得与实际信道环境相匹配的优质数据集是一个非常重要的课题。

4) **AI 模型设计思路较单一**：现有 AI 模型往往聚焦单一的用例和单一物理层功能，在多模态模型设计（例如包括信道估计、预测、压缩等多种用例的 MIMO 信道信息获取模型）、发射机与接收机多模块联合设计、分布式 AI 算法等方面还缺乏高效解决方案。

5) **系统谱效提升难-信道难获取**：精确信道信息对于大规模 MIMO 系统提升谱效至关重要。业界已有多种基于导频的信道估计、预测、压缩相关的 AI 算法，但是很少考虑基于 AI 能力的导频设计、基于信道簇演化特征的信道预测、信道预测和压缩一体化、时分双工（Time Division Duplex, TDD）/频分双工（Frequency Division Duplex, FDD）兼容以及单小区/多小区兼容、模型可在线训练的器件非线性补偿、无真值信道估计等技术方向，这些方向对于提升信道精度以及系统的谱效至关重要。

6) **系统谱效提升难-编码技术提升谱效空间小**：传统信源与信道编码通常基于信息等重要性假设，在任务导向场景中无法适应多模态信源的非均匀特性，产生显著冗余。深度学习驱动的信源信道联合编码可实现任务导向的高效通信，显著提升系统等效谱效，但面对未知或时变信道时仍缺乏泛化与自适应能力。

7) **系统谱效提升难-正交频分复用（Orthogonal Frequency Division Multiplexing, OFDM）系统提升谱效难度大**：循环前缀（Cyclic Prefix, CP）属于冗余开销，在高速移动场景下谱效损耗可高达 25%。现有无 CP 方案在性能、复杂度和灵活性方面仍有明显不足。近年来的深度学习方法虽展现潜力，但在无 CP 场景下鲁棒性和实时性欠佳，且多集中于接收端，缺乏收发机一体化设计，难以应对复杂时频双选信道的需求。

8) **用户调度与 MIMO 预编码算法复杂度高、泛化性差**：随着用户、资源和天线数增加，传统穷举或贪婪算法计算复杂，而启发式方法又导致资源分配效率低。现有学习方法多聚焦窄带系统或单一策略，未能利用策略的置换特性，难以泛化至动态用户场景；部分强化学习方法训练不稳定、收敛慢，且仅学习单一

策略。系统级多小区用户调度在高负载下难以实时掌握全局信道，多小区协作的复杂度问题更严峻。

9) **空口 AI 多模块集成与仿真难度大**：当前空口 AI 研究主要聚焦于单模块性能优化，而端到端的 AI 空口设计复杂度高，目前难以实用。空口性能指标涵盖容量、谱效、能效、算法复杂度及边缘用户干扰消除等多个维度。如何筛选关键 AI 模块有效集成至传输链路，并通过系统级仿真实现多维度性能指标的联合验证，已成为业界亟待突破的核心难题。

2. 白皮书章节安排和主要技术突破

研发团队针对上述诸多挑战开展联合技术攻关，白皮书章节和主要贡献如下：

第三章介绍了 6G 空口 AI 技术设计理念，提出了空口 AI “**能力-质量-代价**” **三角形设计评估模型**。优化目标是在满足 AI 模型时延与泛化性要求的前提下，最大化性能与开销的综合性价比。为实现该目标，白皮书提出两大路径：一是在 AI 模型全生命周期中系统性降低代价，包括通过数据增强减少数据采集开销，以及利用模型轻量化、量化、剪枝等技术降低训练与推理复杂度。二是通过联合优化设计提升效率，例如将信道估计、预测与预编码等多个模块整合为一个 AI 模型，从而在降低总开销的同时提升整体性能。

第四章提出了 **6G 智能内生接入网架构**。传统外挂式、烟囱式的 AI 引入方式无法释放空口 AI 的全部潜力。因此，白皮书提出在接入网层面新增数据面、计算面，并增强控制面与用户面，通过“**分层集中控制、功能分类聚合、柔性按需部署**”的原则，实现连接、计算、数据、AI 模型四要素的深度融合与统一编排。这一架构为空口 AI 任务提供了有服务质量（Quality of Service, QoS）保障的训练与推理服务，是支撑 6G 实现高水平网元智能化的基础能力。

第五章提出了多种**低开销空口 AI 数据集构建方法**。直面 AI 模型训练对高精度标签数据的依赖以及实测数据采集开销巨大的行业痛点，创新性地提出了融合通信知识的新型数据增强框架，将实际数据采集开销降低 80%以上。创新性地采用伪随机算法实现自同步标签生成与事件触发在线训练，有效解决了 AI 接收机训练数据采集成本高与模型泛化性差的挑战。

第六章介绍了空口 AI 传输关键技术方面的进展。面对 5G 导频开销高和场景适配能力差的问题，6G 需向环境感知动态导频演进，本章第一节提出两种新型导频：“疏密导频”通过循环发送密集与稀疏导频，由 AI 模型在时频域依次进行精准估计，动态适配不同移动速度以降低开销；“性能检测导频”则通过少量新增导频来评估估计准确性，并反馈至发送端以动态调整导频间隔，实现资源的高效利用。此外，还介绍了多种 MIMO 信道获取方法。例如**基于簇演化的信道预测方法**。核心创新在于通过“分布对齐”来提升模型泛化能力：首先将复杂信道解耦为独立的簇，再对每个簇进行精细对齐到一个统一的中间表示，从而消除环境变化导致的分布偏移。基于此框架，衍生出信道压缩反馈、时域预测和频域外推三大低开销、高精度应用。该方案在仿真到现实的测试中展现出卓越的泛化能力和性能优势，为实际部署提供了有效解决方案。

在信道获取的基础上，第二节首先提出**单小区用户调度与 MIMO 预编码的联合智能优化方案**：针对大规模多用户 MIMO 系统中组合爆炸和干扰协调的难题，构建了基于双图神经网络（Graph Neural Network, GNN）的联合学习框架，将复杂的联合优化问题分解为调度与预编码两个紧密耦合的嵌套策略。该方案不仅学习性能接近最优数值算法，更具备了卓越的维度泛化能力，可自适应动态变化的用户数与资源块数，推理复杂度相比传统算法降低数个数量级。在多小区场景下，进一步引入了基于多智能体强化学习和知识图谱的干扰协调机制，实现了系统级吞吐量与公平性的全局优化。

第三节提出了一种**具有无速率特性的信源信道联合编解码方案**，核心思想是构建一种基于多项式-高斯分布的新型变分自编码器，能够根据链路状态调整编码开销实现无速率传输，有效应对动态变化的链路条件。同时，在解码端引入通信任务导向机制，使解码器能够根据任务需求自适应地调整特征聚合策略，实现任务驱动的高效信息重建。此外，该方案的无速率特性使得信源信道编码模块能够分别部署在信源侧和空口侧，便于联合编码方案应用于移动通信网络。

第四节面向时频双选信道，提出一种 AI 辅助的 OFDM 脉冲成形与信号检测联合设计方法，抑制信道时频扩展造成的自干扰，提高多载波信号检测精度和谱效。在发射端，采用深度神经网络实现脉冲成形滤波，有效抑制符号间与载波间干扰；在接收端，设计基于卷积神经网络的均衡解调模块，充分利用时变信道相

关性，提升解调性能并降低导频、循环前缀开销。此外，结合图神经网络和元学习，改进的 OFDM 信号检测网络采用离线训练与在线微调相结合的方式，实现对不同场景的快速适应，具备良好的鲁棒性与泛化性。

第七章设计了一套空口 AI 集成与系统级仿真方案，将通过 AI 方法获得的链路-系统平台接口曲线与多个 AI 模块整合至统一仿真平台，全面评估空口多维性能。该方案集成三个映射表，包括发射端信道误差模型、接收端信道误差模型、非线性补偿增益，以及单小区/多小区调度与 MIMO 预编码模块两个集成模块，并联合验证了系统容量、谱效、能效、复杂度、边缘干扰消除等多个系统核心指标。

第八章介绍了集中式与分布式大规模 MIMO 样机以及超高阶调制样机，外场实测数据表明，所提出的 MIMO 信道压缩反馈、非线性补偿等技术方案，在真实信道环境下均取得了显著性能增益，为空口 AI 技术的产业化落地提供了坚实的实践支撑。

第九章是白皮书的总结。

3. 6G 空口 AI 技术设计理念

业界对空口 AI 的设计评估准则还没有达成共识，很多研究还是侧重于大算力前提下的通信能力最大化，对引入 AI 技术后所带来的数据、算力、复杂度等代价以及 AI 模型的泛化性、推理时延等缺乏系统性思考和优化。作为 5G 和 6G 通信系统智能化演进的方向指引，科学完善的空口 AI 设计评估准则不仅是技术创新的基础保障，更是通信产业可持续发展的先决条件。在这一章，白皮书将重点介绍 6G 空口 AI 的均衡设计理念。

3.1. 设计思路

为了更加全面的评估智能空口的性能，中国移动提出空口 AI 能力-质量-代价三角形均衡设计评估模型。如下图 3-1 所示，空口 AI 能力包括引入 AI 技术后谱效、能效、系统容量等空口通信能力指标；空口 AI 代价代表为了达到预期的空口传输能力所付出的 AI 代价，包括 AI 模型相关的数据采集、数据集构建、

模型训练、模型分发等所需的算力、复杂度、空口时频资源等开销；空口 AI 质量是 AI 模型的质量，包括推理准确度、推理时延、不同场景下的泛化性等。



图 3-1 空口 AI 能力-质量-代价三角形均衡设计评估模型

空口 AI 能力、质量与代价的联合优化是一个高度复杂的多目标优化问题。为了简化问题，对其进行了合理的拆解。首先，AI 模型的推理准确度直接决定了空口 AI 能力。其次，推理时延是一个硬性约束，所有可选的 AI 模型都必须满足时延要求，这可以通过优化代价（算力与模型推理复杂度）来实现。因此，在满足推理时延的前提下，单部署场景中的能力-质量-代价均衡优化可以通过最大化能力与代价的比值来实现。

当考虑多部署场景时，AI 质量中的泛化性成为关键指标。泛化性的本质是模型在未见数据上的表现，通常可以通过偏差和方差来评估。单部署场景中的均衡优化可通过最大化能力与代价的比值实现；而在多部署场景中，在满足泛化能力的前提下，能力-质量-代价的均衡优化可通过最大化多场景中能力与代价的综合比值（M）来实现。例如，当空口 AI 能力为谱效时，M 的一种表达式如下：

$$M = \frac{1}{N} \sum_{i=1}^N \frac{w_i SE_i}{C_i}$$

其中，N 是场景数， SE_i 是第 i 个场景中 AI 技术的谱效相比非 AI 技术谱效的增益（倍数），需要大于 1， w_i 是场景的加权， C_i 是第 i 个场景中 AI 技术的综合代价（包含算力、复杂度、空口资源等）与非 AI 技术综合代价的比值。综合代价的定义可以有不同方式，包含综合成本（计算多种代价对应的总成本）以及多种代价的加权和等。

M 最大化的设计准则适用于空口 AI 单模块设计，例如信道估计、预测、MIMO 预编码、MIMO 检测等。该准则同样适用于一个 AI 模型实现多个串行模

块功能的情况，例如 TDD 系统中的上行信道估计、信道预测、MIMO 预编码等功能。

如果空口 AI 有多个能力目标，比如既有谱效又有能效目标，就需要设计一个综合的能力指标，例如谱效和能效的加权。由于系统的能效和谱效单位不同，无法直接相加，所以也可采用增益（倍数）进行相加的方式，如下式所示。

$$M = \frac{1}{N} \sum_{i=1}^N \frac{w_i(EE_i + SE_i)}{C_i}$$

其中， EE_i 和 SE_i 是第 i 个场景 AI 技术带来的能效和谱效各自的增益（倍数）。值得注意的是，空口 AI 相关的数据采集、模型训练、推理等所需的功率要纳入计算能效时系统的总体功耗里面。如果代价中的空口资源体现到谱效，复杂度体现到能效，综合代价可以仅由算力开销衡量。

3.2. 优化策略

为了在能力、质量和代价间实现联合优化，提出两种方法：第一种是在确保空口 AI 能力和质量满足需求的前提下，通过降低空口 AI 的代价来提升 M ；第二种是在空口 AI 能力满足需求的基础上，通过代价和质量的联合优化来提升 M 。

针对第一种方法，可以在 AI 模型生命周期的每个环节采取有效措施以降低代价。具体而言：1) 在数据采集阶段，可以在少量实测数据的基础上进行数据增强，从而减少数据采集的开销。当需要反馈采集数据时，可以仅反馈实测数据的特征，以降低反馈开销。2) 在模型训练阶段，采用更先进的 AI 模型和训练方法，以降低训练复杂度。同时，使用特征更全面的数据集来提升模型的泛化性，减少再训练的需求。3) 在模型再训练阶段，采用低训练开销的在线训练方法，降低模型再训练的复杂度。通过轻量化的迁移学习或元学习，提升 AI 模型对通信场景和参数的泛化能力，从而降低再训练的复杂度。4) 在模型部署和推理阶段，通过低精度量化降低推理的复杂度，并利用模型蒸馏和剪枝技术减少模型参数规模。

对于第二种方法，关键在于确保模型推理的准确度满足通信能力提升的要求，同时推理时延要符合空口用例的时延要求。在此基础上，主要挑战是如何在提升模型泛化性的同时，不增加甚至降低代价。这需要更加巧妙的 AI 模型设计，具备良好的泛化性、低训练/再训练开销、低数据开销和低推理复杂度。

3.2.1. 降低空口 AI 代价

3.2.1.1. 降低数据采集和数据集构建的开销

构建基于少量实测数据的低开销数据集对于提升 M 至关重要，主要原因如下：1) 通信场景和信道环境具有高度的多样性和复杂性，仅依赖仿真数据难以与实际环境完全匹配，从而影响 AI 模型的性能表现。2) 获取实测数据集涉及数据测量、存储和反馈等环节，开销巨大（例如终端测量下行信道数据并反馈给基站），在实际应用中面临较大挑战。为降低实测数据的获取成本，可以在少量实测数据的基础上进行数据增强。中国移动在前期研究中，基于实测信道数据提取了信道在多径时延、角度、功率等方面的统计特征，并通过仿真平台生成了与多种实际信道环境高度吻合的信道数据集。仿真结果表明，该方法能够将实际数据采集开销降低 80%以上，同时将数据反馈开销减少 95%以上。

3.2.1.2. 降低模型训练/再训练和推理复杂度

在 AI 领域，已有多种通过低复杂度神经网络设计来降低模型训练和推理复杂度的方法，这些方法能够显著提升 M 值。例如，采用低精度量化技术可以降低推理复杂度，而剪枝技术则能够有效减少模型的参数规模，从而提升整体效率。

此外，随着通信场景和系统参数的变化，AI 模型的再训练往往伴随着较高的开销。因此，降低再训练复杂度也是提升 M 值的关键手段之一。具体有两种主要方式：一是采用低训练开销的在线训练方法。例如，在接收端通过神经网络对发射与接收器件的非线性特征进行补偿时，由于非线性特征变化较为缓慢，可以选择轻量化的 AI 模型，并基于极小量的训练数据进行在线训练。二是通过轻量化的迁移学习或元学习技术，提升 AI 模型对通信场景和参数的泛化能力。此外，多任务学习可以在不同场景下选择最优的 AI 模型，从而避免再训练。如果希望在降低推理复杂度的同时提升模型的泛化性，可以采用低复杂度的图神经网络设计来实现。这些是降低开销和提升性能的有效解决方案。

3.2.2. 空口 AI 单模块能力-质量-代价联合优化

以 MIMO 信道压缩反馈为例，针对 AI 算法泛化性不足的问题，提出了一种

综合考虑终端和基站算力、基站与终端模型交互开销、模型在线训练开销以及多场景泛化性能的单编码器多解码器 AI 架构。在该架构中，终端侧的编码器在不同通信场景中保持不变，其训练数据集涵盖所有场景的信道数据；而解码器则根据具体场景进行针对性设计，并能够根据场景动态切换。这一设计显著降低了对终端算力和存储的需求，减少了空口信令开销，同时提升了在多场景下的性能。与基于多场景数据集单独训练的多编码器多解码器架构相比，所提架构在仅需 1/5 编码器存储开销就可实现相同的信道获取精度。这种 AI 模型设计结合前文所述的低开销信道数据集构建方法，可实现能力-质量-代价的联合优化。

3.2.3. 空口 AI 多模块能力-质量-代价联合优化

以传统 TDD 系统的大规模 MIMO 信道获取和预编码为例，当天线端口数从 32 增加到 128 时，相应的导频开销占比会增至原来的 4 倍，这对系统谱效的提升极为不利。如果为了保持导频总开销不变而降低每个端口的导频密度，则会导致信道估计误差增加，进而使 MIMO 预编码性能下降。当信道估计和 MIMO 预编码两个模块独立设计时，按照固定导频密度进行配置，天线端口数的增加无法让谱效等比例提升，反而可能导致谱效恶化，造成严重的资源浪费。

在 TDD 系统中，可以通过 AI 模型实现上行信道估计、信道预测以及下行 MIMO 预编码等多个模块的联合设计。具体而言，神经网络以多用户信道信息作为输入，输出未来某一时刻每个用户的 MIMO 预编码矩阵，设计目标为最大化多用户和速率。与采用独立 AI 模型分别实现信道估计、预测和 MIMO 预编码的传统方式相比，多模块联合 AI 设计能够同时降低空口 AI 代价、提升能力并改善质量。主要原因如下：1) 训练数据的采集开销大幅减少，模型训练和推理的复杂度显著下降，从而有效降低了空口 AI 的代价；2) 得益于 AI 技术在识别干扰、噪声特征以及信道相关性等方面的强大能力，即使在低导频开销下，也能实现精确的信道估计和预测。随着天线端口数的增加，导频开销不会等比例增长，甚至可能下降，从而显著提升了空口 AI 的能力；3) 随着端口数的增加，传统预编码算法所需的矩阵运算复杂度显著上升，而在算力受限的条件下，AI 驱动的预编码算法能够在低复杂度的前提下实现比传统算法更高的系统容量，从而进一步提升了空口 AI 的能力；4) 由于 MIMO 预编码复杂度的降低，推理时延也得以减

少，从而有效提升了空口 AI 的质量。这些优势共同推动了方案 M 值的提升。

3.2.4. 空口 AI 模型选择方法

空口 AI 模型在实际部署中面临三重挑战。首先，模型训练复杂度居高不下，表现为样本复杂度、空间复杂度和时间复杂度均处于高位，难以满足对低功耗、低时延的严苛要求。其次，场景泛化能力不足，模型普遍缺乏对信道分布等环境参数的泛化能力，以及对用户数等系统规模变化的维度泛化能力，难以适应无线环境与时变业务带来的动态变化。最后，可解释性较差，从其他领域直接迁移的 AI 模型缺乏对无线空口特性的适配，导致模型选择严重依赖经验性试错。

针对这些挑战，现有研究存在明显局限。为降低复杂度而采用的模型轻量化技术，往往以牺牲性能为代价。在场景泛化方面，研究多聚焦于参数泛化，而对更具挑战性的维度泛化探索不足。在可解释性方面，现有方法要么采用深度展开技术使推理复杂度显著增加，要么依赖事后解释难以直接指导模型的选型与优化。为此，本白皮书提出系统化的设计思路，从学习范式和模型结构两个维度构建空口 AI 模型的科学选型框架。AI 模型的选择应紧密结合空口任务的数学特性：

1) 监督学习：适用于具备明确标注数据的任务，如信道估计与预测、符号检测等闭集识别问题。

2) 无监督学习：适用于需要自主发现数据内在规律的任务，如用户接入调度、功率与资源块分配、预编码设计等优化问题。

3) 强化学习：适用于具有序列决策特性的场景，可建模为马尔可夫决策过程的任务，如移动性管理与切换优化。

模型结构的选择需充分考虑空口策略的数学结构与数据统计特性：

1) 全连接神经网络：具备通用函数逼近能力，适用于缺乏显式结构先验的空口策略建模。

2) 卷积神经网络（Convolutional Neural Network, CNN）：基于局部连接与平移不变性假设，天然契合信道估计等具有空间相关性的数据处理。

3) 循环神经网络（Recurrent Neural Network, RNN）：通过隐状态传递捕获时序依赖性，适用于信道预测等时间序列建模任务。

4) Transformer 架构：通过自注意力机制建立全局依赖关系，特别适合信道

估计、预测与压缩等需要空时频联合建模的场景。

5) GNN: 基于图拓扑结构与置换不变性设计, 可有效处理多用户通信、分布式资源分配等图结构优化问题。

3.2.5. 适配空口的低复杂度、强泛化性、可解释 AI 模型设计

在设计空口 AI 模型时, 充分利用无线策略中固有的结构先验 (如广泛存在的置换对称性) 是实现低复杂度、强泛化性和高可解释性的关键。基于置换对称性的深度神经网络设计能够显著降低模型复杂度。通过将置换对称性融入模型结构, 可以大幅缩减假设空间, 从而有效降低训练复杂度。当策略满足置换对称性时, 单个训练样本可通过置换操作生成大量等效样本, 极大提升了样本利用效率。同时, 具有置换对称性的深度学习模型天然具备参数共享机制, 能够在保持学习性能的前提下, 显著降低模型的空间复杂度和时间复杂度。置换对称性的运用还能显著增强模型的维度泛化能力。通过精心设计的参数共享机制, 可使深度神经网络的规模与系统维度解耦。当用户数、资源块数等系统规模参数发生变化时, 离线训练的模型无需重新训练或微调即可直接应用于未见过的维度场景, 展现出卓越的维度泛化特性。

此外, 充分利用无线策略的置换对称性和函数结构可有效提升模型的可解释性。这些无线领域的先验知识能够预测不同 AI 模型在各类场景下的学习性能, 为深度神经网络的结构设计提供理论指导, 从而系统化地提升模型性能、可扩展性和维度泛化能力。通过系统化设计具有置换等变特性的深度神经网络 (如 GNN 或 Graph Transformer), 可改变传统的“试错式”设计范式。这种基于问题本质的设计方法从无线通信的数学特性出发, 构建模型的输入特征和架构体系, 确保所设计的深度神经网络严格满足无线策略的置换对称性要求。

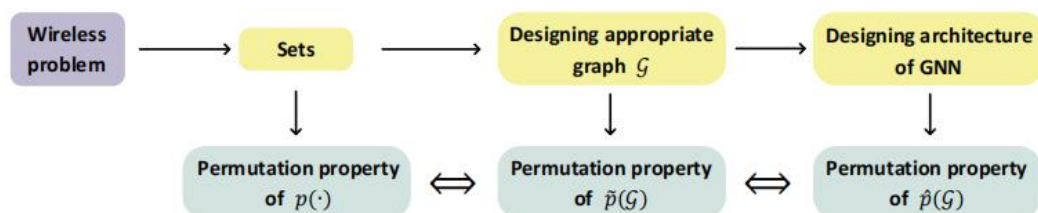


图 3-2 系统化 GNN 设计思路

上图 3-2 以置换等变 GNN 为例, 给出了系统化、通用的设计框架: 首先从

无线问题中寻找集合及集合间的关系，从而确定相应无线策略的置换性质；其次根据集合构造图，即定义顶点/边及其类型和特征，使定义在该图上的函数具有集合导致的置换性质；最后，根据顶点和边的类型设计 GNN 结构，即确定处理与聚合函数的共享方式，使 GNN 表示的函数族满足无线策略的置换性质。

4. 6G 智能内生接入网架构设计

前面已阐述空口 AI 的评估准则，以及能力、质量与代价联合优化的方法。为充分发挥智能空口传输技术在网络中的价值，还需网络与 AI 双向赋能的接入网架构作为支撑。当前接入网架构在支持智能空口传输方面存在不足：一是仅作为数据管道，无法满足 AI 数据全生命周期管理需求；二是网络与计算分离，难以支持 AI 业务的实时性与动态性。因此，需构建统一的 6G 智能内生架构以实现“连接、数据、计算、模型”四要素深度融合。本章节将介绍智能内生接入网架构的驱动力、设计原则和设计方法。

4.1. 6G 智能内生网络架构需求和驱动力

国际电信联盟（International Telecommunication Union, ITU）框架建议书为 6G 创新指明方向，网络与 AI 融合已经成为 ITU 规范的 6G 六大场景之一。AI 是满足移动通信网络新指标的关键路径，赋能网络提升网络运行效率、降低运维成本、增强用户体验。移动通信网络是实现 AI 泛在普惠的基础平台，使能 AI 成为泛在化的社会级服务。为了精准、高效的利用 AI 新生产要素实现生产力变革，以及动态、可靠的供给 AI 新服务能力实现服务范式变革，需要统一的 6G 智能内生智简架构，来满足网络与 AI 融合，实现网络升维。

现有的网络架构在数据全生命周期管理和计算服务性能方面均存在不足。首先，现有网络仅作为数据传输管道，无法满足 6G AI 新型数据要素的采集、传输、处理、存储需求。其次，现有网络和计算相互独立，无法满足 AI 业务实时性、动态性、个性化需求。因此，在“差异大的新增功能集独立成面、功能相近的控制功能全局统一”的总体原则下，网络需新增面向数据的全生命周期管理，面向计算的高性能处理机制两类功能，并增强现有控制面与用户面功能，以 6G 智能

内生架构实现 6G 内生 AI 相关信息服务全流程的支撑。

4.2. 6G 智能内生网络架构总体设计原则

从 ITU 框架建议书的应用场景看，多要素融合是 6G 系统创新的核心特征。多要素融合需要网络能力、网络效率、网络质量的全局最优创新，6G 智能内生架构的设计对实现 6G 全局最优至关重要。面向网络平台化、服务多样化的发展趋势，6G 架构通过连接、数据、计算多要素协同的架构设计，赋能网络内生 AI、通感一体、天地一体等新型网络能力来扩展网络能力，通过低碳、熵减的架构设计，有效提升资源利用效率、网络能效，并降低网络运营成本，通过动态灵活部署、本地闭环自治的分布式自治网络架构设计，有效增强网络可靠性、鲁棒性、安全性来增强网络质量。

在 6G 架构的设计原则上，首先要遵循适用原则，核心功能支持 80% 业务需求，20% 长尾需求可选，以避免过度设计。该原则的关键考虑因素是性能满足需求，快速规模化部署。然后是至简原则，即“简单优于复杂”。如果没有必要，不增加实体，简化结构和逻辑，避免不必要的复杂性，该原则的关键考虑因素是高效低成本，稳定高可用。最后是柔性原则，即“柔性动态优于一步到位”，快速上线新服务的基础是柔性架构，其能够满足各类 AI 任务需求。该原则的关键考虑因素是开放式、服务化架构，灵活可扩展。

基于平台化、服务化的演进方向，白皮书提出“三体”、“四层”分体分层框架，在架构层面实现对新型信息服务全场景、全流程的支持，并结合 6G 场景任务，完成网络本体内各层的要素、层级、连接协同设计。

4.3. 6G 智能内生网络架构设计方法

4.3.1. 场景驱动的网络和 AI 任务设计

任务设计方面，需要明确面向生产型 AI 和服务型 AI 的任务及 AI 业务流程。由场景驱动，研究架构需支持的服务或完成的 AI 任务，基于二八定律筛选主要任务。AI 赋能网络方面，从性能增益（谱效、系统容量）、AI 效率（复杂度、开销）、AI 质量（推理时延、泛化性）等原则出发，筛选出支撑物理层、核心

网的 AI 业务流程，包括终端触发网络侧模型离线训练、基于 AI 的核心网策略优化等。从价值潜力、可行性、技术优势筛选出综合 AI 服务业务流程。网络使能 AI 方面，服务包括 AI 连接服务、AI 计算服务、AI 数据服务以及综合 AI 服务。从实时性、移动性、高隐私保护需求等原则出发，筛选出服务发现、AI 服务的 QoS 评估保障、AI 连接的移动性管理等综合 AI 服务业务流程。

4.3.2. 基于关键流程拆解和分析的要素设计

要素设计需要对关键流程的拆解和分析。AI 赋能网络方面，目标是实现功能的高性能、高可靠和高效能。通过选取 AI 赋能网络典型端到端流程，确定连接、数据、计算、模型相关原子功能要素，提炼出 AI 任务控制、模型训练、模型推理、模型性能监控、模型存储、模型分发、模型选择、数据控制、数据采集、数据处理、数据存储、模型/数据传输移动性保障等主要功能要素。网络使能 AI 方面，目标是实现性能可保障的 AI 服务。通过选取网络使能 AI 典型的端到端流程，确定连接、数据、计算、模型相关原子功能要素，提炼出 AI 业务接入控制、AI 业务编排、AI 任务协同控制、数据控制、模型控制、数据采集、数据处理、数据存储、AI 推理执行、AI 计算执行、用户连接转发功能增强等主要功能要素。

由于每一个流程有独立的要素，因此，由流程驱动，基于奥卡姆剃刀原则和“自包含（原子服务功能完整、独立）、可重用（可被多个服务调用）、独立管理（独立部署、升级、迁移）”原则确定要素形态。具体包括 AI 业务编排、任务控制，两者均更多服务于网络使能 AI，某些情况下，AI 赋能网络也需要任务控制。此外需要增强控制面和用户面的功能，并新增数据控制与数据执行功能、计算控制与计算执行功能，并在流程管理中把模型管理作为重要模块进行提炼。

4.3.3. 基于资源供给能力和复杂环境特征的层级设计

在需求方面，网络使能 AI 需求具有多样性，AI 赋能网络在核心网、接入网也存在共性和差异，例如，物理层的实时性要求高于高层和核心网。在供给方面，从核心网到物理层，算力、数据、连接能力均存在差异。因此，需要研究如何在复杂多变信道、用户移动性和复杂拓扑环境中，合理布局各技术域要素，达到能力、效率和质量的均衡，实现需求和供给的最优匹配。

层级设计基于能力、效率、质量均衡原则，通过统一结构框架和要素按需分布，实现要素供给与任务需求间的最优匹配。在分布式 AI 执行中，考虑时延差异性、隐私要求、传输开销，AI 执行要素的分布原则，将算力按需分布在最需要计算的触发点上。在分层集中式协同控制中，考虑实时性需求，协同控制效率、可靠性要求，AI 控制要素的分布原则，对接入网区域级实时控制，实现通/算/数/智多维资源实时控制调度、多终端/基站节点的集中跨节点协同、实时适配无线空口动态变化。

4.3.4. 基于柔性原则的高效至简的连接设计

连接设计方面，6G 内生 AI 将引入计算、数据和模型相关新资源和新功能，连接复杂度增加，由服务化设计理念出发，基于柔性原则，设计高效至简的连接关系和接口协议。

5G 的核心网引入了服务化接口，实现了控制和管理功能的集中处理和接口服务化，摒弃了固定的控制连接和刚性流程，实现了灵活的服务化调度和柔性服务编排。虽然 5G 核心网原子化的逻辑功能相比 4G 有了很大程度的增加，但是服务化接口的引入大幅降低了接口连接设计的复杂度。6G 内生 AI 架构的设计充分考虑到这一设计理念，将新引入的多要素协同的功能进行集中化处理和接口服务化；此外新引入的业务执行处理相关的功能，比如计算执行功能，数据执行功能，模型管理功能等，引入服务化的接口，可以实现按需在不同层级进行部署，实现资源的复用和灵活的接口调用。

4.4. 6G 智能内生总体网络架构设计框架

通过定任务、定要素、定层级和定连接，白皮书提出 6G 内生 AI 架构方案构想。架构具备三大特征，首先是分层集中控制，由核心网业务集中控制，接入网区域级实时控制。第二个特征是功能分类聚合，数据类功能面向全要素的数据生产和消费，计算类功能面向 AI 计算的算网一体融合控制和执行。最后是柔性按需部署，根据业务场景需求，按需部署分布式执行功能。

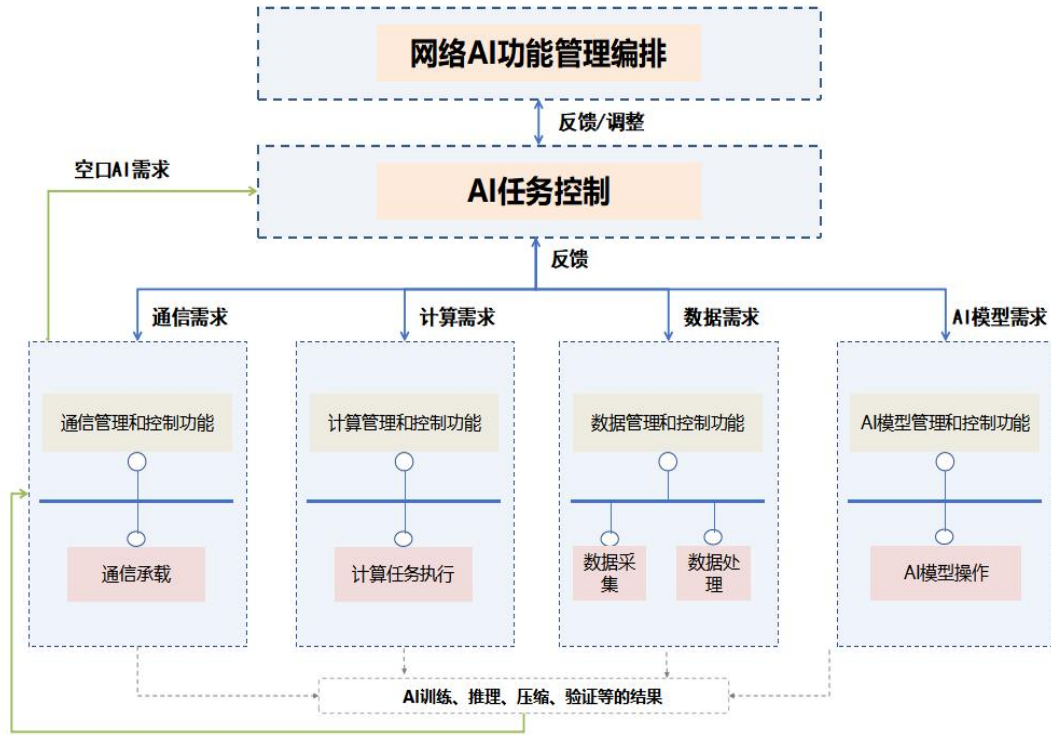


图 4-1 6G AI 内生架构设计框架

基于 AI 内生的系统架构设计，白皮书提出新增数据面、新增计算面、增强控制面、增强用户面，完善服务化功能层设计。

其中，面向 6G 智能内生的网络架构总体设计由三级构成：网络 AI 管理编排管理、AI 任务控制和通信/计算/数据/AI 模型管理和控制功能。

网络 AI 管理编排管理：实现切片级的 AI 需求处理。不同的 AI 服务可能需要不同的网络功能，如 AI 训练服务需要训练所需数据采集、数据处理、连接建立、计算资源控制、AI 训练等网络功能，而 AI 推理服务需要连接建立、计算资源控制、AI 推理等网络功能，不同网络功能之间的连接拓扑也可能不同。因此，当收到新的 AI 服务请求且当前网络已有服务功能链无法满足时，网络 AI 管理编排管理功能对网络功能进行编排。

AI 任务控制：负责 AI 任务的生命周期管理，并实现用户级的 AI 需求处理和策略下发。不同用户需求（可以是网络设备的某个 AI 需求）输入到 AI 任务控制，AI 任务控制对任务的创建、修改、删除等生命周期进行管理。如果当前编排好的功能服务链可以满足，AI 任务控制对用户请求进行解析，转化为网络可执行的指标，如时延、准确度等。进一步的，AI 任务控制将指标分解为对通信连接、计算、数据、AI 模型的需求，分别下发给相应的网络功能。

通信/计算/数据/AI 模型管理和控制功能：负责 AI 任务执行阶段的控制和调度，包含信息收集和资源控制两个大模块。其中，信息收集是指需要实时感知多节点的算力负载、数据处理能力、当前使用的算法模型、以及通信管道的信道条件。基于上述信息收集，通过更实时的资源控制能力，例如随着网络环境的变化，通过模型、数据、连接及算力的实时调度，进行实时的需求监控和保障。

4.5. 6G 智能内生接入网架构方案设计

6G 智能空口引入 AI 能力（比如 MIMO 信道估计、信道预测、信道压缩、预编码、编码调制、干扰协调等）在大规模商用落地存在较大的挑战，一是效率低，算法复杂度和算力需求高，对于空口数据的采集、存储和处理开销大，尤其是无线侧毫秒级别的高精度信道数据对基站的存储和处理要求高，并且可能还需要终端反馈，占用宝贵的空口资源；二是能力弱，算法跨场景泛化弱，不能匹配网络的大动态变化，无线应用的场景配置和信道环境差异巨大，比如终端不同的移动速度、视距或者非视距、不同信干噪比、不同的用户分布、信道多径分布等，不同的网络配置差异大，比如多种基站类型、多种天线规模、不同的频带等问题。三是质量差，网络节点面临算力受限，难支持实时 AI 的推理，可靠性难保证。

6G 在设计阶段就要考虑和 AI 的深度融合，不同于 5G 通过 AI 功能叠加、外挂等方式，6G 空口 AI 内生在架构层面支持连接、计算、数据和 AI 算法/模型等元素的深度融合，可针对不同场景需求实现以任务为中心的统一灵活编排和控制，按需提供有 QoS 保障的 AI 训练和推理服务，支持将 AI 能力按需编排到无线空口，为实现高水平的网元智能化提供基础能力。

4.5.1. 面向空口的智能内生接入网功能和接口研究

AI 功能无线内生：为了实现效率、能力和质量的均衡需求，在无线接入网元内部增加 AI 推理和在线训练所需的算力、数据执行等逻辑功能，让数据处理、AI 模型推理/在线训练更靠近用户，满足空口 AI 用例对于效率、能力、质量的均衡需求。

AI 无线任务控制：在接入网架构中引入 AI 任务控制功能，负责 AI 任务（比如 AI 模型的离线训练，在线训练，推理，性能监测等）的生命周期管理，实现

连接、计算、数据、模型四要素的调度控制，将 AI 任务控制拆解为对通信连接、计算、数据、AI 模型的 QoS 需求，同时引入 AI 计算和 AI 数据能力的感知功能，实现连接功能、计算功能、数据功能、AI 模型功能模块间的协同，这样可以高效的适配无线网络环境可能的动态大幅变化，在模型性能下降的时候能够快速触发模型的在线训练，满足多场景下的泛化性需求。

新增的无线数据面：目前网络数据的采集大多采用控制面接口，但是随着 AI 的广泛应用，海量数据采集的开销越来越大，控制面的机制很难满足海量数据的采集需求；现有的无线用户面连接用于终端和核心网用户面功能（UPF）进行应用层数据的端到端处理，上述处理难以满足 AI 相关的大数据存储、大数据转发、大数据路由、大数据计算等功能需求，此外缺乏终端与无线接入网元和网元间的数据采集通道和机制，这些迫切需要引入新的数据面的传输机制来进行数据的采集和控制，由新增的无线数据面来负责 AI 模型传输、大数据的存储转发以及大数据清洗、格式转换、AI 训练中间数据的传输等功能。

增强的无线控制面：6G 无线控制面在无线资源控制（比如无线连接承载的会话，移动性功能）之外，还要引入 AI 计算和 AI 数据相关的会话和移动性控制功能，比如 AI 模型控制、计算控制以及数据控制等功能，底层的物理层传输协议可以采用同一套协议，对高层空口协议做进一步的增强。

服务化和非服务化融合接口：智能内生无线接入网的逻辑功能的接口还面临着服务化或者非服务化的选择。一方面，6G 业务需要快速、按需部署，以适配更为广泛和多样化的 6G 应用场景。同时，人工智能技术及应用有其独立的发展路径，客观上需要 6G 系统中的智能功能可以独立演进、迭代部署、即插即用，服务化的逻辑功能可以较好地满足这些需求。另一方面，6G 智能内生无线接入网需要重点关注实时、严格时序的无线处理，非服务化的逻辑功能在这一方面更有优势。因此，对于新增功能，对于时延要求不高、独立演进要求较高的功能引入服务化的接口，对于时延要求苛刻的功能沿用现有的非服务化接口。

端侧智能：随着终端的算力和存储功能越来越强，在终端引入数据处理、AI 模型推理甚至在线训练等功能也成为可能。无线网络中为了 AI 计算存在不同类型的终端，比如提供 AI 训练数据的终端，提供 AI 计算的终端，提供 AI 服务请求的终端。在网络的协调下，不同类型的终端与网络侧的计算、数据和 AI 功

能协同，利用海量终端的规模优势，一起提供 AI 的模型训练和推理服务。

4.5.2. 6G 智能内生接入网架构设计建议

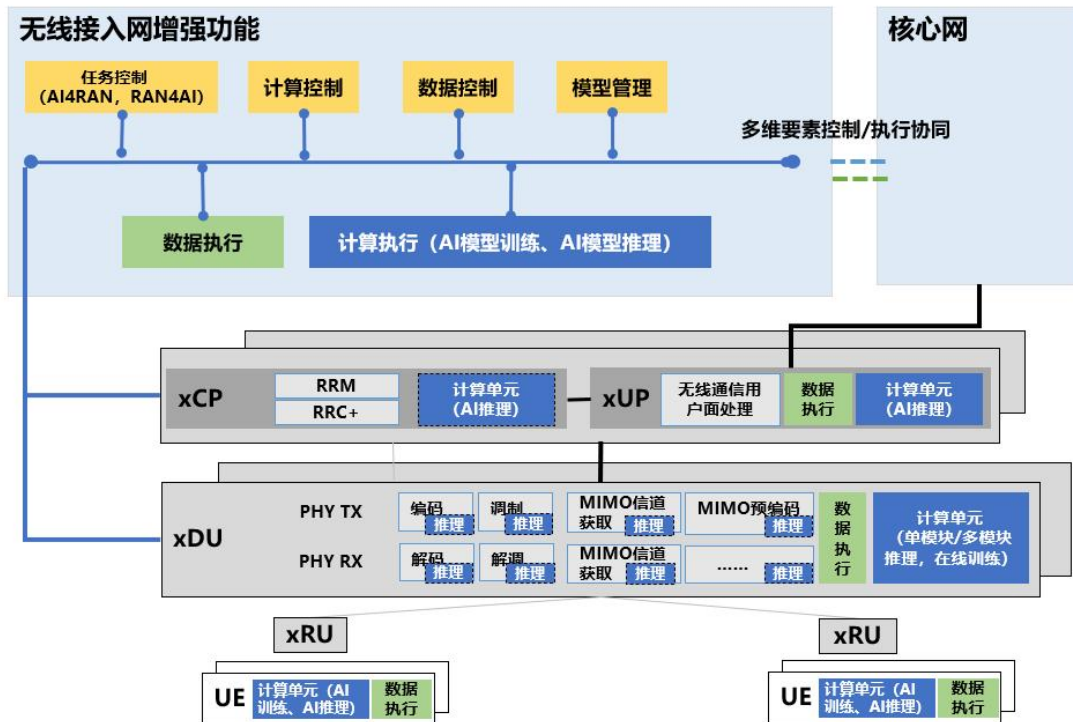


图 4-2 6G 智能内生接入网架构建议

6G 智能内生接入网架构建议如图 4-2 所示，接入网分为已有功能的增强和新增功能，具体的功能描述如下：

xCP: 6G 基站集中式单元-控制面，负责 L3 控制；引入计算单元进行 AI 模型的推理，比如基于 AI 的移动性管理模型的推理等；引入无线资源控制（Radio Resource Control, RRC）功能的增强，用于与数据控制、计算控制等功能进行交互，进行数据和计算相关的会话控制。

xUP: 6G 基站集中式单元-用户面，负责无线通信的用户面处理，引入计算单元进行用户面 AI 模型的推理，此外引入数据执行功能，引入数据执行的功能进行新增数据面所需的数据传输所需的转发功能。

xDU: 6G 基站分布式单元（Distributed Unit, DU），主要调度、物理层收发机（比如编码、调制、信道获取、预编码等）等无线处理功能，引入计算单元使能单模块或者多模块 AI 功能的推理，甚至在线训练。

无线接入网新增逻辑功能，新增逻辑功能与已有增强功能之间采用服务化接

口进行连接。1) 任务控制：负责 AI 任务的拆解和控制。2) 计算控制：负责计算功能的控制，负责 AI 计算资源的感知和控制等。3) 模型管理：负责 AI 模型的监测、选择等功能。4) 数据执行：负责无线数据收集、存储和处理功能，5) 计算执行：负责 AI 模型的训练、推理等功能。

4.6. CIS-RAN：智能内生接入网架构的演进

在上述智能内生接入网架构设计的基础之上，6G 无线接入网还需向协作化和服务化方向发展。基于 2010 年提出的 C-RAN 架构，中国移动进一步提出智能协作无线接入网架构（Cooperative Intelligent Service-based RAN, CIS-RAN）。

在协作化方面，聚焦 6G 网络在性能跃升和能力拓展方面的高需求，CIS-RAN 架构在强化现有协作通信能力的同时，创新性引入协作感知与协作智能功能。

- **支持以用户为中心的协作通信：**针对 C-RAN 固定协作边界导致的性能受限问题，提出以用户为中心的网络化协作 MIMO，实现从单点接入到“Always-on”协作接入、从非相干传输到相干联合传输、从有界到无界三大根本性转变，保障用户一致性体验。
- **支持拓展的协作感知与协作 AI：**为支撑 6G 多样化应用场景，CIS-RAN 创新性引入协作感知与协作 AI，提出在核心网或 RAN 侧部署感知功能节点，兼顾复杂度和灵活性；原创提出 RAN 集中 AI 单元，构建集中-分布相结合的智能处理机制，实现了通信、感知、AI 数据的智能高效处理。

在服务化方面，5G 核心网通过平台化与服务化设计，已实现了从固定资源供给向灵活业务供给的范式变革。随着 6G 对业务定制化与差异化需求的进一步凸显，无线接入网也需考虑向服务化方向演进。然而，从技术实现角度看，RAN 在构建松耦合、可复用且独立管理的原子服务方面，较核心网面临更大挑战。因此，建议 RAN 的服务化演进路径可从“能力可开放”起步，逐步向“能力可定制”推进。

- **RAN 能力可开放：**针对 RAN 能力通过核心网间接对外开放，开放能力受限、时延大问题，提出通感算智可开放的 RAN 服务以及意图驱动的

RAN 能力开放技术，实现更高效的对外赋能。

- **RAN 能力可定制：**针对差异化、定制化业务需求难满足的问题，提出无线链路控制功能分段下沉、HARQ-ARQ 重传功能合并等功能模块化重构技术，以及服务自生成、多接口协议灵活选择等技术，实现功能高内聚设计与场景化按需组合，敏捷适配场景的定制化功能与资源需求。

5. 低开销空口 AI 数据集构建

当前，空口 AI 技术的规模化部署面临数据采集开销巨大的核心瓶颈。传统方法依赖海量实测数据，导致采集、存储与传输成本高昂，同时存在模型泛化性不足的严峻挑战，因此构建低开销、高效率的数据集不仅是技术突破的焦点，更是产业生态成熟的重要前提。本章将系统性阐述在信道状态信息（Channel State Information, CSI）压缩反馈、室内定位及 AI 接收机等典型场景下，如何通过知识融合、稀疏采样与自同步标签等创新方法，实现数据的低开销采集与模型性能的显著提升。

5.1. CSI 压缩反馈低开销采集方法

5.1.1. 业界面临的挑战

信道数据涉及高维、跨域的特性，其结构涵盖接收端口域、发送端口域、时间域与延迟域等不同信号维度，每一维度均呈现独特的统计分布特征。此类结构复杂性使得传统数据增强方法难以有效挖掘信道数据中的物理特征与内在模式，因而在无线通信任务中表现不佳。

为应对该挑战，已有研究探索面向通信信号特性的专用数据增强手段，如引入了信号旋转、翻转和高斯噪声注入等增强操作，并将其整合于深度学习模型的训练与推理流程中。尽管取得一定进展，但现有数据增强技术仍存在显著不足：多数方法缺乏对信道物理结构与传播机理的深度融合，难以在多维信号域中保持数据结构一致性；同时，它们通常依赖于大量实测数据作为增强基础，未充分利用通信领域知识与信道建模成果，导致数据利用率低、泛化能力有限。

5.1.2. 思路和方法

为克服上述局限，白皮书提出一种融合通信知识与 AI 建模的新型数据增强框架，其核心区别在于将信道物理特征与统计模型引入数据生成过程，从而实现数据增强。该方法的主要流程包括：

1) 场景化实测数据采集：在目标区域开展定点实测，获取典型下行链路信道响应数据。通过有限但具代表性的实测样本，提取该场景下的关键传播特征。

2) 信道参数提取与统计分析：基于实测数据，提取多径传播参数及环境相关参数，包括时延扩展、角度扩展、路径损耗等，并采用数学统计工具建立各参数的分布模型，形成对传播环境的量化描述。

3) 仿真平台参数校正：将统计获得的信道特征参数进行扩展使其具有更多的特征，并嵌入 3GPP 等标准仿真平台生成信道数据，从而构建与实测环境高度匹配的数据集。

4) 场景精细化划分与数据库构建：基于所提取的信道统计参数，对标准宏观场景（如城市宏蜂窝（Urban Macrocell, UMa））进行精细化划分，构建多个子场景。利用大量实测与增强后的仿真数据，为每个子场景预先构建专用的信道数据集与预训练 AI 模型库。

5) 动态场景匹配：当用户进入一个新场景时，首先进行少量实时测量并提取信道关键统计参数；随后，将该参数集与预构建的子场景数据库进行快速匹配，识别出最适用的子场景类别；最终，直接调用该子场景对应的数据集与预训练模型。

5.1.3. 成效与先进性

所提出的通信知识联合 AI 模型的数据增强方法，在实践中展现出显著优势：

1) 实现高效的数据采集与特征提取：通过所提方案，在实际场景中选择少量的代表性位置进行信道测量，与传统数据采集方法相比，样本采集开销降低了 90%以上。

2) 具有卓越的模型泛化性能：所提场景适配信道模型（Scenario Specific Channel model, SSCM）进行数据增强的方法在测试集（模型未见过的信道）上展现了优秀的性能和泛化能力，如下图 5-1 所示，SSCM 生成的数据（红色曲线）

性能显著优于传统 3GPP 模型（黄线）、标准码本（蓝线）及传统实测与加噪方法（绿线）。

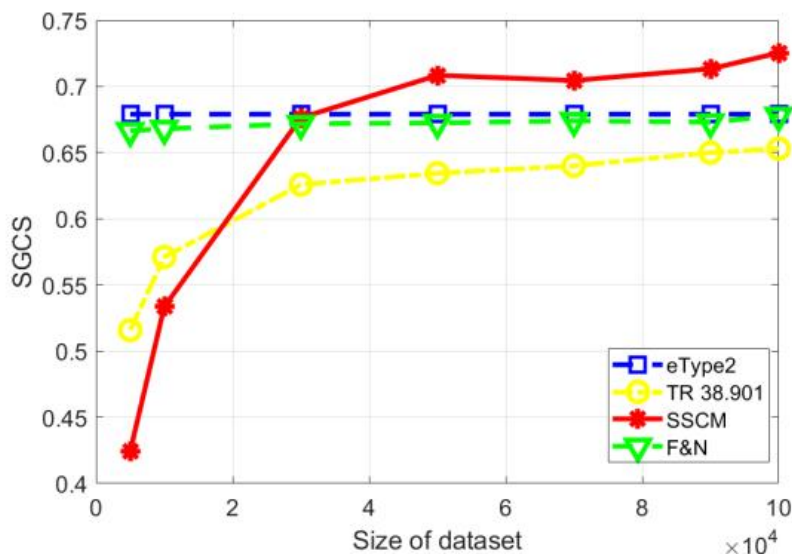


图 5-1 不同样本数性能对比图

3) 实现超低的空口反馈开销：所提方法仅需上传 20 个 32 位精度的信道统计参数（640 比特），相较于传统方案需上传 5 万个原始信道矩阵所耗费的 5120 万比特，空口反馈开销的降低超过 95%，极大缓解了上行链路的容量压力，为 AI 模型的大规模部署扫清了关键障碍。

4) 实现高效的场景迁移与快速部署：通过将新部署场景 D 的信道参数与预划分的子场景库（如 A、B、C）进行匹配，实验表明直接采用匹配子场景 B 的数据集，其性能与基于场景 D 实测数据训练的结果高度接近（如表 5-1 所示）。该方法在几乎不损失精度的前提下，显著降低了数据采集与模型训练的时间成本，有效支持 AI 模型在动态环境中的快速部署。

表 5-1 场景参数对比

Dataset	A	B	C	D
μ_{lgDS}	-7.6	-6.8	-6.0	-6.6
σ_{lgDS}	0.7	0.675	0.65	0.66
μ_{lgASD}	1.26	0.7	1.6	0.75
σ_{lgASD}	0.3	0.25	0.28	0.24
μ_{KF}	10	8	7	8.3
σ_{KF}	4	3	4	2.8
SGCS性能	0.8655	0.9127	0.8775	0.9133

5.2. 低开销 AI 室内定位数据采集

5.2.1. 业界面临的挑战

基于 AI 的室内高精度定位技术是 6G 泛在连接的关键环节，但其规模化部署面临严峻的数据采集瓶颈。该技术的核心在于构建映射信道特征与物理位置的“指纹库”模型，而这一过程面临多重挑战：首先，完整的信道数据包含多基站、多天线端口、多时延路径等高维特性，常表现为海量数据，这为前传链路和存储系统带来了显著负担；其次，为训练高精度模型，传统方法需要进行密集的网格化实地测量，采集数万至数十万个人工标注样本，耗费的人力与时间成本极高，严重制约了系统的快速部署；此外，利用信道模型生成的仿真数据因理想化假设而与真实环境存在系统性偏差，直接用于训练会导致模型性能恶化。数据维度高、采集效率低与仿真真实性不足，共同推高了系统总体成本，严重制约了该技术在工业物联网、智慧仓储等场景中的规模化商用。

5.2.2. 思路和方法

为应对上述挑战，白皮书提出了一种低开销 AI 室内定位数据采集与处理方法。该方法的核心思路是通过降维及数据增强技术，在最大限度保留定位信息的前提下，显著降低对实测数据的依赖。具体方法如下：

1) 数据维度精简：在非视距（Non-Line of Sight, NLOS）占主导的室内工厂场景中，通过摒弃传统使用全部天线端口数据的做法，转而采用单端口数据进行模型训练。可采用两种端口信息融合策略：一是简单的功率平均；二是利用 AI 模型训练出的最优加权平均，从而保留关键信道特征。

2) 样本数量优化：信道在空间上具有连续性，相邻位置的信道特征具有高度相似性。利用这一原理，无需对所有位置进行密集实测。通过采用网格稀疏采样策略，仅采集少量关键位置的实测数据；然后，通过线性插值或 K 近邻插值等算法，根据已知采样点的数据，拟合生成大量新采样点的信道数据，从而低成本地构建出覆盖全面的定位指纹库。

3) 数据来源拓展: 构建带偏见的半监督学习框架, 该框架通过教师模型同时预测位置和输出表征置信度的偏差, 进而指导学生模型在训练中对海量、无标签的仿真数据进行置信度加权学习, 自动削弱低质量仿真数据的影响。

5.2.3. 成效与先进性

通过系统性地实施上述方案, 在保证定位精度的同时, 实现了数据采集开销的显著降低:

1) 维度精简: 采用单端口加权平均方案后, 基于信道冲激响应 (Channel Impulse Response, CIR) 的定位精度仅从 0.64 米略微下降至 0.73 米, 远优于 3GPP 规定的 1 米定位精度要求。与此同时, 数据集的存储开销直接降低至原来的 1/4, 极大地缓解了数据存储与传输压力。

2) 样本优化: 通过稀疏采样与插值扩充, 仅需采集 6400 个实测样本 (占全量样本的 22%), 即可通过插值生成与 28800 个全量实测样本相当的数据集, 并实现 1 米的定位精度。数据采集开销降低了约 70%, 大幅缩短了现场部署时间。

3) 性能提升: 带偏见的半监督学习方法 (红色线) 性能优异, 如图 5-2 所示, 在相同实测数据量下, 其定位精度显著优于传统监督学习和半监督学习方法。此外, 在达到相同定位精度的同时, 该方法所需的实测数据量比传统监督学习减少约 40%, 实现了数据利用率的最大化。

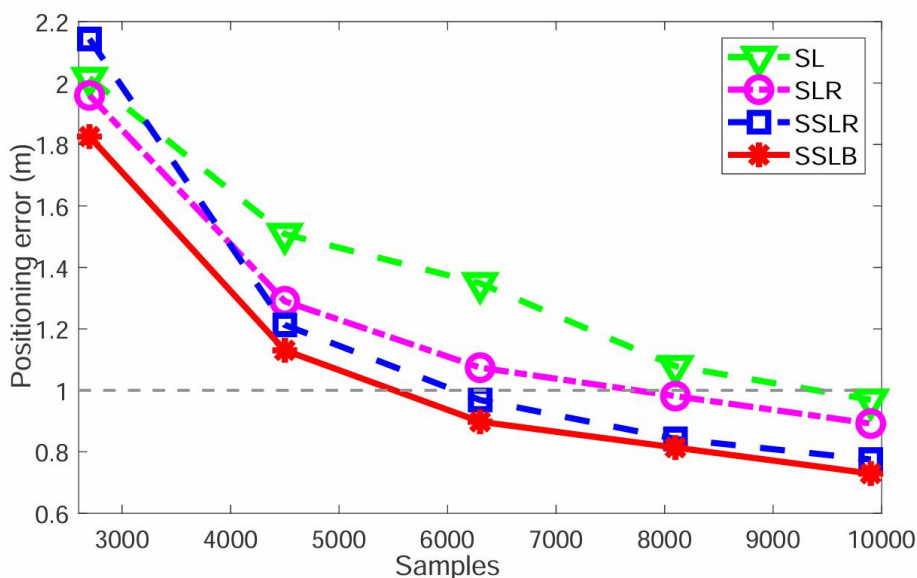


图 5-2 仿真结果对比

5.3. AI 接收机低开销数据采集

5.3.1. 业界面临的挑战

AI 技术为高性能接收机设计引入了全新范式，但面临两大核心挑战：首先，监督标签空口采集开销巨大。AI 接收机训练依赖“发射比特流-接收信号”的配对数据。仅以单时隙、单资源块（Resource Block, RB）、正交相移键控（Quadrature Phase Shift Keying, QPSK）调制为例，10 万样本即需约 33.6 兆比特标签。实际网络中，随着资源块、调制阶数（如 1024QAM）的提升，数据量将呈指数级增长，导致难以承受的采集、存储与回传开销。其次，模型泛化能力严重不足。基于特定数据集训练的接收机模型，其性能高度依赖于训练时的信道环境与器件特性。一旦部署到差异较大的新场景中（如信道变化），可能导致解调性能急剧恶化。

5.3.2. 思路和方法

为应对上述挑战，白皮书提出了两种创新的自同步标签生成的解决方案，将标签数据的“采集”转变为“本地生成”，从而实现低成本的数据采集。

1) 基于伪随机算法的自同步标签生成机制：摒弃传统方法中必须从发射机端物理采集比特流作为标签的做法，本文利用通信双方预先同步的伪随机生成机制，在本地生成完全一致的随机序列作为训练标签。具体而言，在通信初始阶段，基站与终端同步一套伪随机数生成算法及其初始种子与配置参数。当接收端解调性能下降明显，需要进行模型再训练时，基站指示终端在指定时频资源上发送训练数据流；与此同时，基站与终端基于已同步的算法，在本地独立生成相同的伪随机比特序列。终端将该序列调制后发送，而基站则直接将本地生成的序列作为 AI 模型训练的监督标签，启动在线训练，优化模型在新场景下的性能。

2) 基于导频序列的自同步标签生成机制：为了最大化降低系统信令开销和

终端复杂度，实现“无缝”集成，可直接复用现有通信标准中的导频序列生成机制。网络侧仅需通过极简的信令（如 1 比特标识）通知终端是否采用自同步标签生成机制，并配置再训练的时频资源与调制编码方案。终端无需集成多套算法或处理复杂的信令，只需像生成导频一样生成训练序列即可，极大地简化了流程，降低了终端侧的资源消耗与管理复杂度。

5.3.3. 成效与先进性

所提面向 AI 接收机的低开销数据采集展现出显著的成效与技术先进性：

1) 通过自同步标签生成机制实现数据采集开销与解调性能的均衡：该机制将训练数据的标签采集成本从吉比特级大幅降低至近乎零，消除了从发射机端采集和传输海量比特流标签的巨大开销。同时接收机能够灵活启动在线训练，使 AI 接收机可持续适应环境变化，增强了在未知场景下的模型泛化能力。虽然会引入额外的训练数据空口传输开销，但自同步标签生成机制能精确产生训练标签，有效提升模型再训练性能。通过在传输用户数据的同时合理控制训练数据样本数量，可实现空口传输开销与性能提升之间的均衡。

2) 系统信令与终端开销最小化：通过复用导频序列生成机制，避免了引入额外的随机序列算法种子等信令开销，同时简化了终端的操作流程，降低了计算与存储资源消耗，使得方案易于部署。

6. 空口 AI 传输技术设计

基于前面的三角均衡设计评估准则，我们重点攻关空口传输的关键技术：6.1 节介绍了多种新型导频设计与信道估计、预测、压缩反馈方法，6.2 节介绍了单小区和多小区的用户调度与 MIMO 预编码方法，6.3 节介绍了高谱效空口 AI 编译码和波形设计方法，6.4 节介绍了空口 AI 多模块联合设计方法。

6.1. 信道信息获取方法

现有导频设计面临发送端难以动态优化导频图案、接收端难以实时判断信道估计精度的挑战，而基于 AI 的信道获取算法普遍存在环境和维度泛化性不足、器件非线性补偿能力有限等问题，这些因素严重制约了 6G 系统性能全面提升。作为实现高精度、低开销、强鲁棒信道状态获取的核心路径，构建新型导频结构、新型信道估计、预测、压缩反馈方法与智能非线性补偿机制已成为空口智能化演进的关键任务。本章将系统阐述在 6G AI 空口信道信息获取方面的系列创新方法。

6.1.1. 新型导频设计方法

6.1.1.1. 行业面临的挑战

5G NR Release 15 为新空口引入了“按需出现、场景可配”的导频框架，以取代 4G LTE 中始终在线的设计，从而支撑大规模 MIMO 等新需求。其核心是将导频功能细分为解调参考信号（Demodulation Reference Signal, DMRS）、信道状态信息参考信号（Channel State Information Reference Signal, CSI-RS）、相位跟踪参考信号（Phase Tracking Reference Signal, PT-RS）和探测参考信号（Sounding Reference Signal, SRS）四类正交信号，各司其职。然而，该正交导频设计存在如下严重挑战：

1) 发送端难以动态优化导频图案：增加导频以提升精度会挤占数据资源，使系统陷入“估计精度与数据吞吐量”难以兼得的困境。多样化场景中导频的动态优化更加挑战。

2) 接收端难以实时判断信道估计精度：接收机包括信道估计、MIMO 检测、解调译码等全链条多个环节，接收端无法基于译码结果实时判断信道估计是否足够准确，这直接造成接收机无法判断现有导频配置是否与当前场景匹配，影响了发送端动态优化导频选择。

6.1.1.2. 思路与方法

为了从根本上解决上述单一导频模式在复杂通信场景下面临的“精度”与“开

销”难以兼顾的矛盾，以及难以动态优化调整导频配置的难点，白皮书提出了感知移动速度、信道估计准确度等因素的新型导频设计，具体包括疏密导频设计和性能检测导频设计两个例子。

1) 疏密导频设计 (a+b 型)

新型导频结构由两部分组成：a 型密集型导频与 b 型稀疏型导频，如下图所示。两者配合使用，发送方式采用“a + N*b”的重复形式。参数 N 的取值与具体场景相关：例如在高铁等高速场景（速度 > 300 km/h）中，N 取 1-2；而在低速场景（速度 < 30 km/h）中，N 通常大于 10。

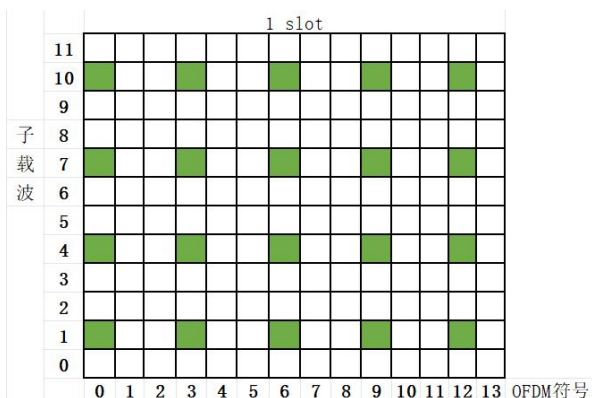


图 6-1 a 密集型导频

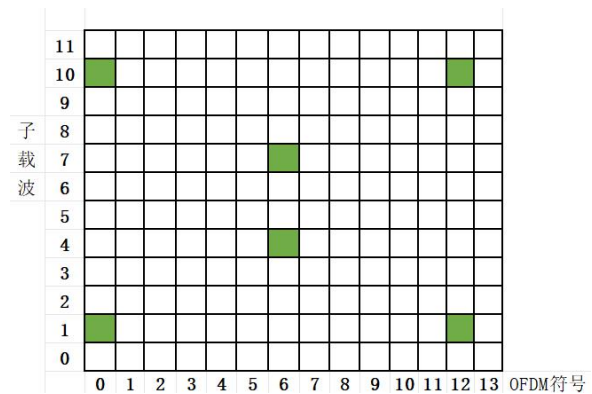


图 6-2 b 稀疏型导频

发送端首先将 a 型和 b 型导频对应的信道估计模型发送至接收端；在确定 N 值后，按“a + N*b”的顺序循环发送导频。在一个“a + N*b”周期内，接收端在接收到 a 型导频后，将其输入基于密集导频的信道估计模型，完成该时刻的信道估计，得到估计结果 H(1)；接收端在接收到第一个 b 型导频后，将其与 H(1) 共同输入基于稀疏导频的信道估计模型，得到估计结果 H(2)。类似地，第 n 个 b 型导频的信道估计仅依赖于前一次估计结果 H(n-1)及当前接收的稀疏导频，从而得到 H(n)。

通过采用上述新型导频结构，接收端可借助不同的 AI 模型，依次在密集与稀疏导频上实现高精度信道估计。该方法不仅显著降低了导频开销，还能够灵活适配多种通信场景。

2) 性能检测导频设计

依赖传统导频收端无法衡量信道估计结果的准确程度，这是业界存在的一大难题。白皮书提出了信道估计性能检测导频，在传统导频基础之上，增加少量新型检测导频（如下图 6-3 所示），收端基于性能检测导频与相邻传统导频的信道估计情况，协助发端判断是否需要调整导频间隔。该导频设计可在很低导频开销情况下实现对信道估计结果的高效准确检测；通过反馈机制，发端可据此灵活且合理地调整导频间隔。

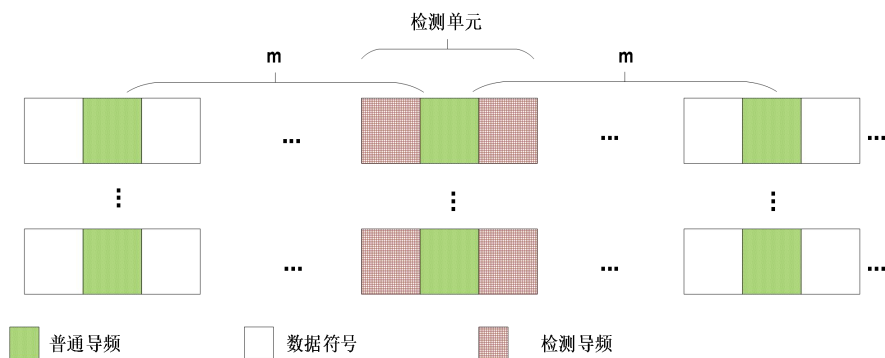


图 6-3 检测单元

6.1.1.3. 成效与先进性

基于稀疏导频，例如在 36×36 的 OFDM 时频网络上每隔 5 个 OFDM 符号插入 1 个导频，白皮书提出了基于双向长短期记忆网络（Long Short-Term Memory, LSTM）+注意力机制的 Seq2Seq 信道估计方案。该信道估计 AI 模型包括编码器和解码器两部分，其中，编码器通过双向 LSTM 提取导频序列的时域特征；解码器通过 LSTM 重构完整信道响应，结合多头注意力机制（Multi-head Attention）增强全局依赖建模。双向 LSTM 能够同时从前向与后向分析信道序列，综合过去与未来的信道状态信息，从而更完整地建模时域相关性与多径效应，尤其在高速场景下精准跟踪信道快速变化。注意力机制则动态聚焦于输入序列中与当前估计最相关的关键部分，自适应关注影响信道状态的核心时频资源块，有效抑制频率选择性衰落与干扰，显著提升估计精度与鲁棒性。该架构可实现低开销、高精

度、低复杂度的 OFDM 信道估计。仿真表明，如下图所示，在 30km/s+20dB 场景下，稀疏资源块上信道估计的平均归一化均方误差（Normalized Mean Squared Error, NMSE）可达到 0.009。

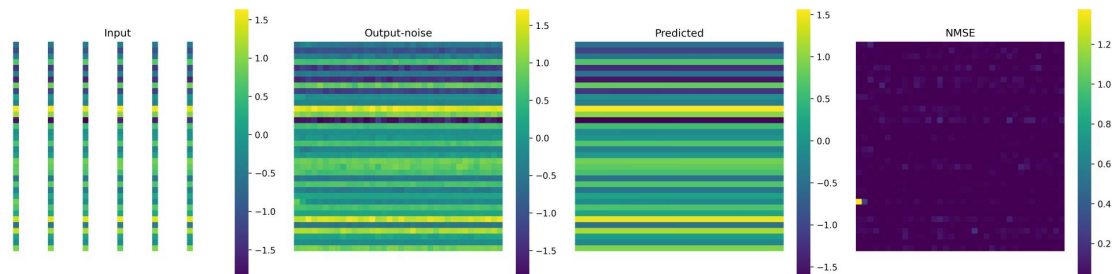


图 6-4 Seq2Seq 信道估计方案的 NMSE 结果

6.1.2. 基于多维度协同的信道压缩

为实现智能准确获取实时信道状态信息 CSI 的核心目标，需围绕多维度技术要素开展协同优化，具体涵盖：CSI 反馈的时效性保障、CSI 获取过程中的上行开销控制、多用户场景下的信道特征相关性挖掘、AI 驱动 CSI 处理模型的灵活可扩展性，以及 TDD 与 FDD 双制式下信道特性的差异性适配与共性利用。

6.1.2.1. 压缩预测一体化反馈

高速移动场景下，终端反馈的下行 CSI 存在因时效性差引发的调度不准确、系统容量受限问题。传统手段为了增加反馈频率，同步增大 CSI 反馈开销，反而会降低系统容量。白皮书提出了压缩预测一体化的 CSI 反馈方案，终端先依据下行导频信号完成下行 CSI 估计，结合当前及历史估计结果，通过三层 ConvLSTM 模型预测下一时刻下行 CSI；随后基于 Transformer 模型，对预测的下行 CSI 进行压缩后反馈至基站。仿真验证，相较于基于 3GPP Rel-16 eType II 码本的 CSI 反馈方案、仅依托 AI 技术的 CSI 压缩方案，该方案的平方广义余弦相似度（Squared Generalized Cosine Similarity, SGCS）分别提升约 7.06% 和 3.15%，能够在不增加反馈开销的前提下，有效抑制高速移动场景中的信道老化影响，同时保障较高的 CSI 反馈精度。

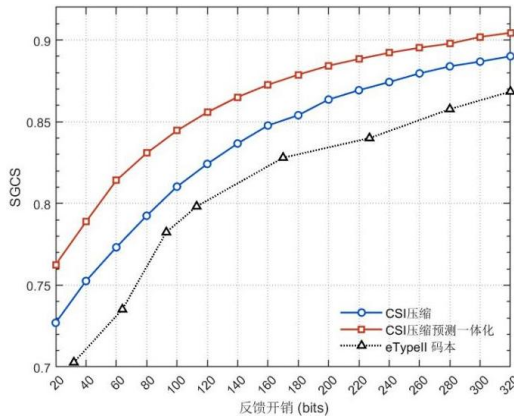


图 6-5 压缩预测一体化方案的 SGCS 结果比较

6.1.2.2. 灵活可扩展 CSI 反馈方案

传统智能化 CSI 反馈方案的机器学习模型，在训练阶段已将特定反馈开销、天线端口数等参数作为约束嵌入架构。当参数变化时，通常需重新训练，严重制约场景适用性；若为不同参数分别训练对应模型，会导致模型数量剧增，给存储、部署与传递带来极大不便，显著增加应用的成本与复杂度。当前主流思路是采用截断机制实现不同的反馈开销，采用补零的方式应对不同端口数量、子带/带宽扩展性问题，但易因截断丢信息、补零引冗余噪声，导致整体性能欠佳。

白皮书提出了一种灵活可扩展的智能化 CSI 反馈方案 SCsiNet，其核心编译码模块在内侧和外侧拥有多条支路，每条支路连接着一个扩展增强模块，其中下/上采样块被引入用于支持多种 CSI 反馈开销，线性变换模块被引入用于维度降低和维度扩张。仿真表明，提出的 SCsiNet 模型在系统内外部条件变化的情况下，无需重新训练，且性能接近每一种对比方案的性能。与基于 Rel-16 eType II 码本的 CSI 反馈相比，在相同开销下，本方案吞吐量可提高 3%~11%。

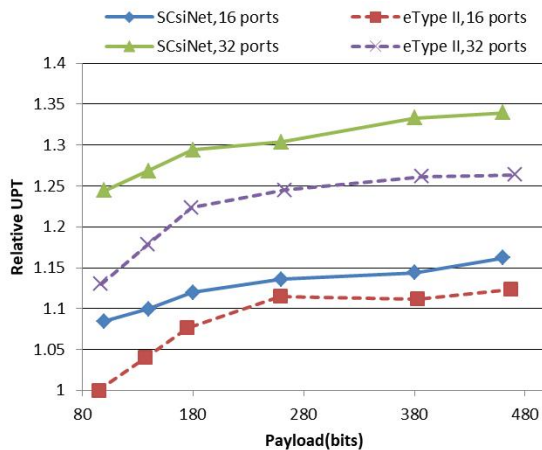


图 6-6 SCsiNet 在 Rank 自适应下的评估结果

6.1.2.3. 多用户信道状态信息融合技术

现有 CSI 压缩算法与 AI 模型往往忽略了相邻用户间存在的强相关性。因为空间邻近性和传播环境相似性，多用户 CSI 往往呈现显著的相关性。通过用户间共享信息进行联合重建，可在不增加额外开销的前提下，有效提升 CSI 恢复性能。白皮书提出通过多头注意力网络 CoTransNet 学习用户共享 CSI 的潜在表征，实现相邻用户间互补信道相关性的动态捕捉，将提取的共享信息与各用户本地特征融合，实现多用户 CSI 的协同重建。仿真表明，相较于传统独立反馈模式，所提方案在全部反馈比特开销下的 SGCS 平均提升 4.98%，NMSE 降低约 3.54 dB。

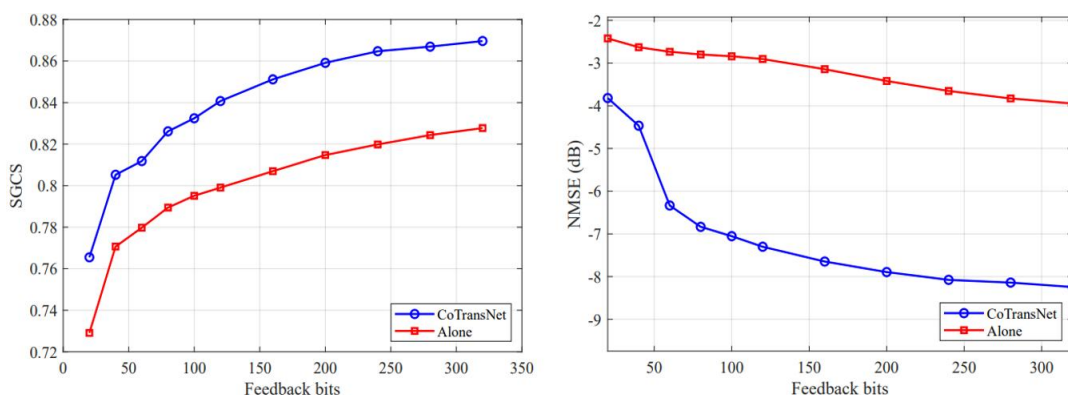


图 6-7 CoTransNet 与独立模型在不同反馈比特数下的性能对比

6.1.2.4. 统一的 FDD 和 TDD CSI 反馈方案

现有 TDD 与 FDD 系统的下行 CSI 获取方案存在设计割裂：TDD 系统依赖上下行信道互易性，而 FDD 系统需终端主动反馈下行 CSI。当通信设备需同时支持双模时，独立的两套算法体系显著增加了硬件成本与复杂度。然而，实际场景中二者存在协同优化空间：FDD 系统的上行与下行信道虽频段分离，但仍存在潜在相关性，而 TDD 系统在高速移动场景下信道互易性也会变差。

白皮书提出了统一的 TDD、FDD 反馈框架，首先基站通过终端发送的上行参考信号获取上行 CSI，然后基于上行 CSI 预测下行 CSI。基于 Conv-LSTM 模型，在 FDD 数据上单独训练主预测模块，然后冻结主模块参数训练 TDD 的预处理模块。仿真结果表明，相较于传统反馈方案，统一反馈模型实现了 3.4%-6.8% 的 SGCS 增益，NMSE 有 0.24dB-0.35dB 的增益。

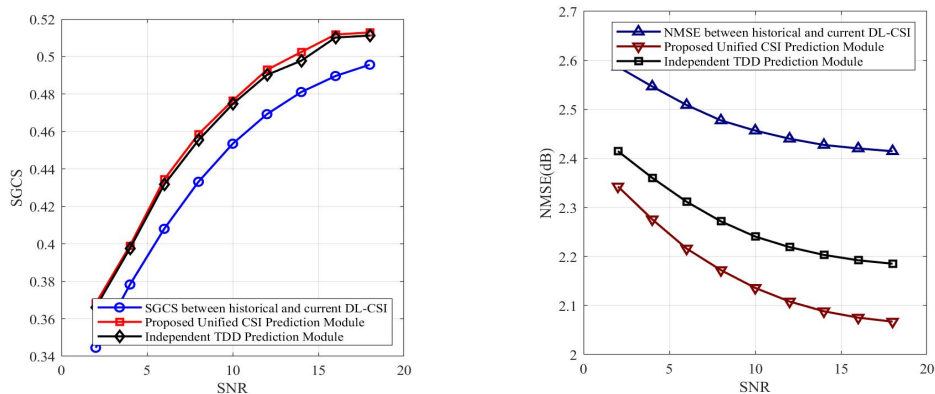


图 6-8 统一 TDD、FDD CSI 反馈方案中 SGCS 与 NMSE 比较

6.1.3. 基于簇演化的信道估计、预测与压缩

6.1.3.1. 行业面临的挑战

随着大规模 MIMO 天线规模持续扩大，系统性能提升正面临新的瓶颈。现有信道获取环节面临双重挑战：一方面，为保障估计精度需要配置高密度导频，导致系统开销增加；另一方面，在受限的开销下，传统方法难以实现高精度的信道估计。这一根本矛盾使得高精度、低开销的信道获取成为释放大规模 MIMO 系统性能潜力的关键技术突破口。

当前主流信道获取方案主要围绕信道预测、频域外推与压缩反馈三大技术路径展开。业界普遍采用 CNN、LSTM、Transformer 等深度神经网络进行优化建模。然而，这些方法仍存在明显局限：现有预测机制在导频稀疏场景下性能显著恶化；基于码本的反馈方案对信道稀疏特性及端口间相关性的利用不够充分。尤其是，尽管深度学习方案在特定场景下展现出优于传统方法的精度与开销平衡能力，但其普遍存在的泛化不足的问题。当部署环境与训练数据分布出现差异时，模型性能会出现急剧下降，这极大限制了其在复杂多变场景中的实用价值与部署效率。

6.1.3.2. 思路和方法

白皮书提出一种基于簇物理过程的可泛化大规模 MIMO 信道获取框架，旨在利用深度学习模型提升信道压缩反馈、时域预测及频域外推的精度，并显著降低获取开销。创新在于通过物理启发的分布对齐方法，确保预训练模型在分布外

新场景中仍保持优异泛化能力，大幅降低实际部署成本。本方案包含三大组成部分：信道分布偏移建模、物理启发的信道分布对齐框架，以及下游信道获取应用。

6.1.3.2.1. 信道分布偏移模型建模

基于机器学习域泛化理论，模型的泛化性由训练阶段与测试阶段数据分布的偏移所决定。基于广泛采用的多簇信道模型，可将大规模 MIMO 信道的分布偏移主要归结为两类：多簇结构分布偏移与单簇响应分布偏移。多簇结构分布偏移：指信道中散射簇的数量以及簇间依赖关系随环境的变化。例如，空旷场景下簇数目较少，而富散射环境中簇数目增多。同时，簇间并非独立，其依赖关系深受环境中物体布局的影响，这是导致分布偏移的关键因素之一。

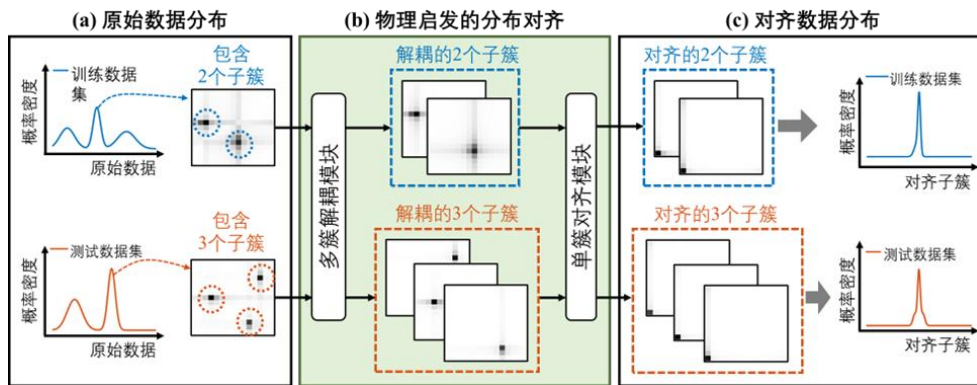


图 6-9 物理启发的信道分布对齐框架

6.1.3.2.2. 物理启发的信道分布对齐框架

为从根本上提升 AI 模型的泛化性，本白皮书提出了一个物理启发的两阶段分布对齐框架。该框架可降低训练阶段与测试阶段深度神经网络输入分布的偏移，其包含多簇解耦与细粒度簇对齐两个部分：

1) 多簇解耦：将复杂的多簇信道分解为多个独立单簇响应的叠加。通过解耦，使得后续处理能够专注于每个簇的内部特征，从而在模型训练时避免拟合特定于训练环境的簇数目分布和簇间依赖关系，有效克服多簇结构分布偏移。

2) 细粒度簇对齐：针对单簇响应分布偏移，设计了对齐算子。该算子利用过采样 DFT 码本进行精细的峰值位置搜索与相位量化，将对齐后的簇响应峰值位置固定至参考点，并缓解功率泄露效应。这一过程将不同环境下的单簇响应分布统一到一个中间表示上，显著降低了单簇响应特性变化带来的分布偏移。

6.1.3.2.3. 下游信道获取应用

基于上述分布对齐框架，本白皮书设计了三种核心信道获取方案对应的模型训练以及推理过程，该方案对于多种深度神经网络结构都具有普适性。

1) 可泛化信道压缩反馈方案：在训练阶段，对训练集中的信道数据进行多簇解耦和细粒度对齐，使用对齐后的簇响应数据集训练编码器和解码器网络。损失函数旨在最小化各簇的重构误差，从而学习到各环境间可泛化的簇相应分布特征。在推理阶段，用户端对实测信道执行解耦与对齐，将各簇的压缩特征比特及必要的元数据反馈至基站。基站端先解码出对齐后的簇响应，再通过特征恢复算子结合元数据，重构出完整的天线-频率域信道。反馈开销与解耦出的簇数目成正比，实现了自适应的开销与精度权衡。

2) 可泛化信道预测学习方案：该方案采用面向簇的神经网络设计，专注于学习单个簇随时间的演化规律，而非整个多簇信道的联合演化，从而规避多簇结构偏移。通过引入循环移位操作，将不同时刻的簇响应与初始时刻对齐，使网络能够学习簇特征的相对变化量，增强了模型对初始状态分布变化的鲁棒性。结合数据增强技术，进一步提升了模型在动态场景下的预测泛化能力。

3) 可泛化信道频域外推学习方案：在模型训练前，首先需从已测频带和目标频带的信道中分别提取簇响应。通过聚类和协变换处理，确保已测频带与目标频带对应的簇响应在特征空间中对齐并保持正确的依赖关系，形成训练数据集。之后，将使用对齐后的已测频带径响应作为输入，目标频带径响应作为标签，训练深度学习模型。推理时，将已测频带的新径响应对齐后输入训练好的模型，预测目标频带的对齐径响应，再经协补偿和叠加，最终外推出完整的宽带信道。

6.1.3.3. 成效与先进性

为了验证所提物理启发的信道分布对齐框架，使用开源 WAIR-D 数据集和 Wasserstein-1 距离定量评估信道对齐效果。从下图 6-10 中可看到所提物理启发的信道分布对齐可降低环境间分布偏移 70%，验证了所提方案的有效性。

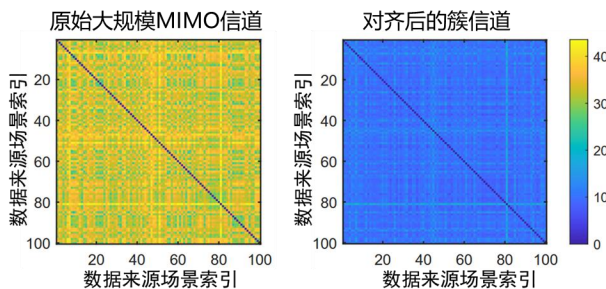


图 6-10 物理启发的信道分布对齐定量评估

本白皮书也考虑 Sim-to-Real 实验验证所提方法在实际信道中的泛化性。其中预训练数据集为 3GPP 38.901 规范下的室内用户信道数据，测试数据集为 Argos 开源户外实测信道数据，包含 4 个 LOS 区域与 5 个 NLOS 区域。完成预训练后，所提方案在不同反馈开销下的信道重构 NMSE 如下图 6-11 所示。可看到，所提方案在实测信道环境下仍能实现低开销，高精度的信道反馈，显著优于对比基线。

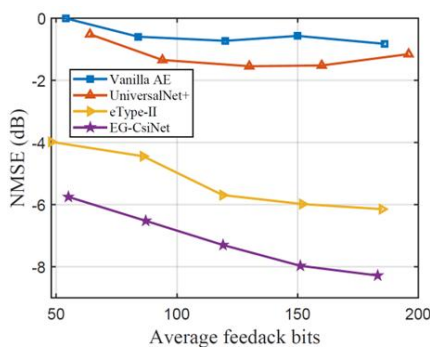


图 6-11 Sim-to-Real 信道反馈实验

6.1.4. 信道非线性补偿

6.1.4.1. 行业面临的挑战

无线通信系统中高阶调制的应用面临显著挑战。随着调制阶数升高，星座图上符号点间距缩小，导致解调所需信噪比门限提高。为维持相同的小区覆盖范围，发射端需增加功率。但通信设备的发射功率受限于功率放大器，一旦输出功率接近或超过其饱和区，信号便会因放大器固有的射频非线性特性而产生失真。

传统上，系统依赖发射端的数字预失真或功率回退技术，以及接收端的基于模型的非线性均衡来抑制此类失真。传统方法存在局限：模型难以精确匹配所有功率放大器的复杂非线性特性，算法实现复杂度高，且在真实信道条件下性能往往不理想。即便是经典的 Volterra 均衡等接收机非线性补偿技术，也因模型普适

性差和接收机难以预知具体放大器特性等问题，在 4G/5G 系统中未能广泛应用。

6.1.4.2. 思路和方法

为了更好的解决非线性干扰问题，白皮书设计了一种智能非线性补偿（AI-enabled nonlinearity compensation, AI-NC）方法，能够在不增加额外开销的条件下实现快速线上训练，可以适应于不同类型的功率放大器特性、多径信道环境，以及多天线多用户场景。上述非线性补偿接收机的结构如下图所示：

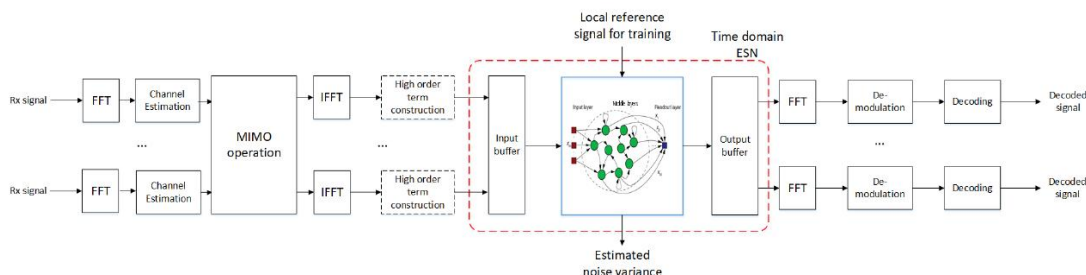


图 6-12 AI-NC 接收机

回声状态网络（Echo State Network, ESN）是 AI-NC 的核心模块，是一个两层神经网络，在输入层后只有一个隐藏层和一个输出层，在 ESN 当中，隐藏层又被叫做储备池。和所有 RNN 一样，输入层的信号会经过加权矩阵 W_{in} 加权后进入隐藏层，而前一次迭代的隐藏层的输出也会经过加权矩阵 W 重新输入到隐藏层当中，隐藏层的输出经过加权矩阵 W_{out} 的加权之后作为输出层的输出。为了缓解 RNN 训练时的梯度消失和梯度爆炸的问题，在 ESN 网络当中 W_{in} 和 W 的系数是固定的，只在 ESN 初始化的时候随机选取权重即可，所需要训练的只有隐藏层到输出层之间的加权矩阵 W_{out} 。

6.1.4.3. 成效和先进性

下图 6-13 列出了仿真参数为中心频率 3.5GHz, 带宽 50RB, 3GPP 抽头延迟线（Tapped-Delay Line, TDL）-A 信道模型，时延为 30ns，发射天线数为 1 配置下的仿真结果：

当调制方式为 256QAM 时，AI-NC 相较于传统线性接收机的性能增益随误差矢量幅度（Error Vector Magnitude, EVM）升高而增大。具体而言，在 1T4R 配置下，EVM 为 2% 时，AI-NC 的信号噪声比（Signal-to-Noise Ratio, SNR）增益不足 1dB；当 EVM 升至 3.5%，增益扩大至 2dB；若 EVM 进一步提高至 6%

甚至 8%，AI-NC 仍可在 24.4dB 与 27.2dB 的工作点实现 10% 的 BLER，而传统接收机已无法达到该 BLER 目标。若接收天线增至 64 根，即便 EVM 为 2%，AI-NC 仍能带来 3.5dB 的增益。

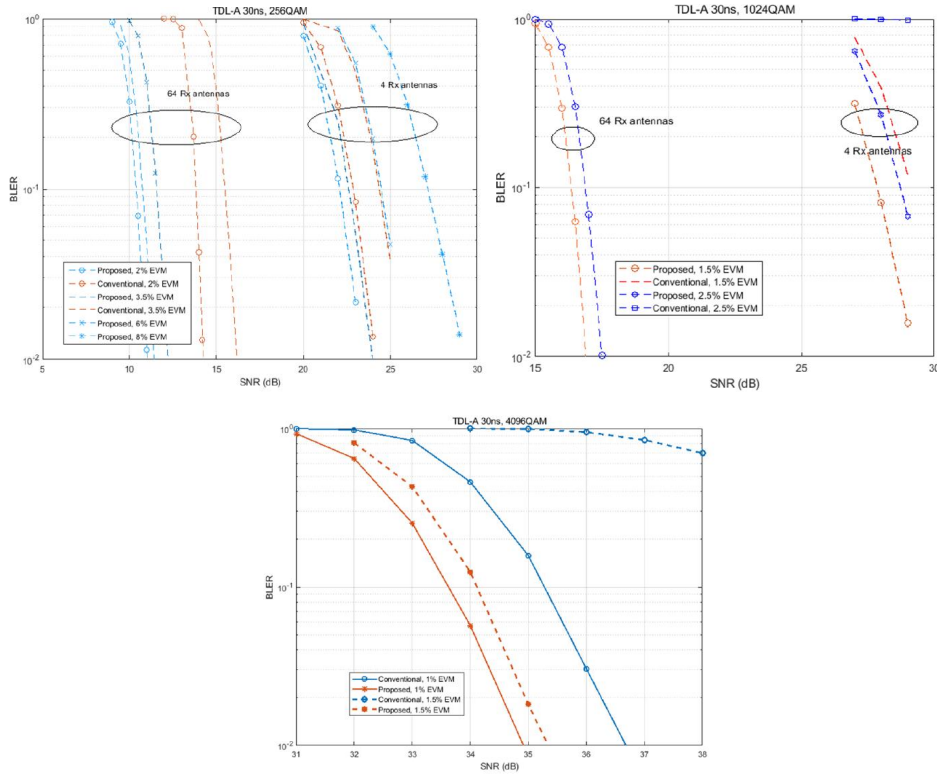


图 6-13 256QAM，1024QAM，4096QAM 的误块率（Block Error Rate, BLER）曲线

在 1024QAM 调制下，若 EVM 为 2.5%，传统接收机无法工作，而 AI-NC 可在合理 SNR 范围内实现低于 10% 的 BLER；当 EVM 为 1.5% 时，AI-NC 的增益为 1dB。若接收天线增至 64 根，传统接收机同样失效，而 AI-NC 仍能保持可靠解调。

对于 4096QAM 调制，在 4 根接收天线条件下，当前 NR 协议尚未支持该调制方式，仿真显示其解调门限为 1% EVM。AI-NC 相较传统接收机有 1.8dB 增益。若 EVM 升至 1.5%，传统接收机已无法解调 4096QAM，而 AI-NC 仍可在合理 SNR 下实现 10% BLER。

此外，在发射天线为 2 的情况下，也对 8 根与 64 根接收天线进行了仿真：

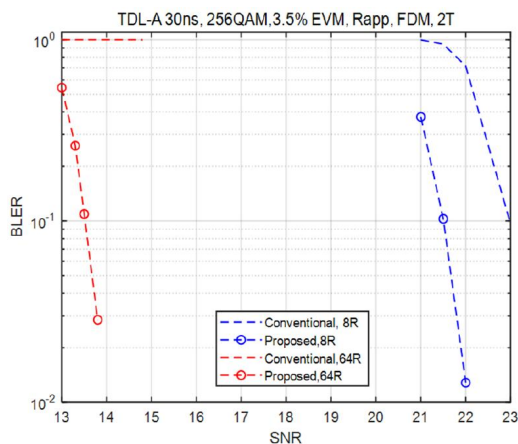


图 6-14 2T&256QAM 的 BLER 曲线

如图 6-14, 在 EVM 为 3.5% 时, AI-NC 在 2T4R 配置下可实现 1.5dB 的 SNR 增益, 而在 2T64R 配置下增益进一步提升至 13.5dB。仿真结果表明, AI-NC 能够在接收端有效补偿发射机的非理想特性, 从而放宽对发射 EVM 的限制, 实现在相同发射功率下传输更高阶调制信号。

6.1.5. 无真值的信道估计

6.1.5.1. 行业面临的挑战

真值信道指在特定时空、频率及极化条件下, 电磁波在收发两端之间传播的完整、精确且无损的信道响应, 涵盖所有导致信号衰落、失真与干扰的物理过程的数学描述。在 AI 驱动的通信系统中, 真值信道作为监督学习的标签, 为模型提供准确目标, 指导其从接收信号中估计信道参数。其数据质量直接决定模型性能, 尤其在复杂动态环境中, 数据需覆盖多样场景来保障泛化能力。

然而, 真实环境中难以获取无噪、精确的信道数值。当前基于导频估计的近似真值用于 AI 训练, 精度与可靠性尚待提高, 同时也会导致模型泛化能力不足。

6.1.5.2. 思路和方法

为了克服现有方法的局限性白皮书设计一种全新且高度可实施的基于 AI 的信道估计框架。其主要创新点包括以下方面:

1) 自监督学习机制: 网络直接在噪声信道观测数据上进行训练, 避免了传统监督学习对干净信道标签的需求, 使模型更适用于无真值的工业实践。

2) 实用的架构部署方法：通过仅替换去噪模块与现有的信道估计集成，确保了 AI 模型可以实现无缝的“即插即用”，有利于 AI 模型的更新和替换。

3) 输入自适应的神经模型：模型可以根据信号维度动态调整计算复杂度，以优化资源利用。该自适用模型更匹配于实际通信系统部署中灵活的参考信号的带宽配置和时域符号配置。

4) 增强鲁棒性的训练策略：通过结合系统不完美建模和噪声增强，所提出的模型可以在实际失真条件下实现鲁棒的性能。

在无真值信道估计中，本白皮书使用一个校准的 NR 平台进行了仿真。仿真涵盖了广泛的低信噪比值（低于 0dB）以及在城市微型信道模型中的实际时偏/频偏条件。动态解调参考信号（DMRS）配置在第 3 和第 12 个符号上，采用 4 个资源块的梳状类型。

通过使用符合 3GPP 标准的仿真器合成噪声信道估计，该仿真器模拟了实际失真。为了处理这些受损信道，设计了一个参数少于 40,000 的紧凑型 U-Net 变体，仅使用卷积层。这种架构通过灵活的滤波操作内在地适应 DMRS 配置，计算成本随资源元素数量呈线性扩展。通过利用卷积的空间归纳偏置，模型高效地捕获无线信道局部时频相关性。

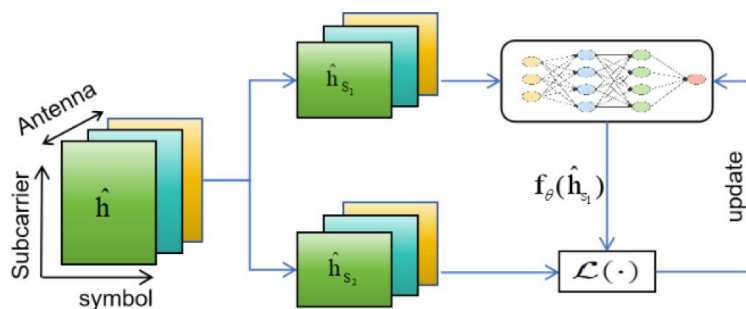


图 6-15 无真值信道估计模型训练方法

训练策略：如上图 6-15 所示，最小二乘（Least Squares, LS）信道估计在频域上被分解，其中数据集 1 包含奇数子载波索引，而数据集 2 包含偶数子载波索引。这种区分是为了利用相干带宽特性，即相邻子载波之间存在强相关性。同时训练双向映射，以实现相互纠错。这种训练范式在实际延迟扩展条件下最大限度地抑制了残余误差。

模型部署：在部署过程中，两个子集的去噪输出沿频轴使用原始的奇偶索引进行重构。应用频域平滑操作以消除训练过程中子载波交错引起的不连续性。这

种两阶段过程保留了原生子载波顺序，产生了更准确反映原始信道的合格信道。

6.1.5.3. 成效和先进性

在无真值信道估计中，为了评估所提出的框架，使用一个校准的 NR 平台进行全面的仿真。这些仿真涵盖了广泛的低信噪比值（低于 0dB）以及在城市微蜂窝（Urban Microcell, UMi）信道模型中的实际时偏/频偏，用户移动速度从 3km/h 到 30km/h，中心频率为 3.9GHz，接收天线数为 16，DMRS 配置在第 3 和第 12 个符号上，采用 4 个资源块的梳状类型。基线方法采用自适应窗口移动平均算法进行去噪，而上界则通过在导频位置考虑理想信道估计来实现。此外，使用与所提出方法相同的训练策略训练了一个有监督的基线，其中基线模型使用干净信道作为标签。

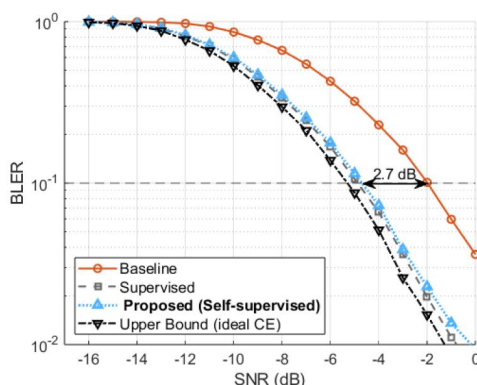


图 6-16 自监督无真值信道估计 BLER 曲线

如上图 6-16，自监督无真值信道估计实现了接近理想性能，在 BLER = 1% 时与上界仅相差 0.3dB，比实际基线高出 2.7dB。值得注意的是，有监督和自监督版本之间的性能差异可以忽略不计（< 0.1dB），证实了所提双向频域学习策略在减轻噪声标签影响方面的有效性。与此同时，所提出的 AI 去噪器展示了实际部署的可行性，仅需约 0.15 MB 的存储空间和 4.9e6 次浮点运算次数（Floating Point Operations, FLOPs）进行推理。

6.1.6. 准多模态大规模 MIMO 信道信息获取模型设计

6.1.6.1. 业界面临的挑战

大规模 MIMO 系统中的信道精确获取对于提升系统性能至关重要。当前业界

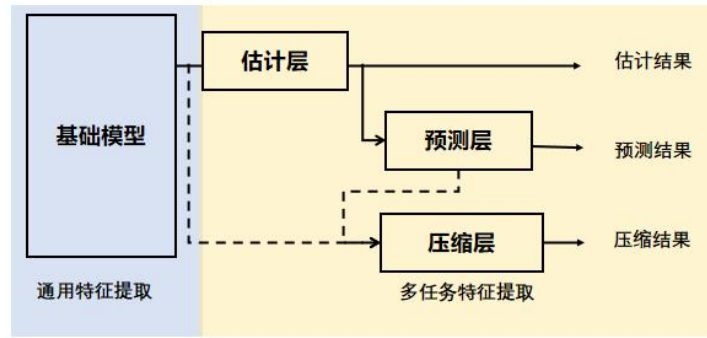
已证明了 AI 模型能够在信道估计、信道预测、信道压缩反馈等多个信道获取用例中可获得可观的性能增益。然而，当前业界普遍采用针对不同用例进行 AI 模型独立训练与推理的方案，随着多个模型被引入系统，其整体的训练、存储及推理计算开销将显著增加，给系统带来较重负担。如何在准确获取信道状态信息的前提下，提高模型的复用能力，降低模型训练、存储和计算开销，成为智能空口技术的一大挑战。

6.1.6.2. 思路、方法与成效

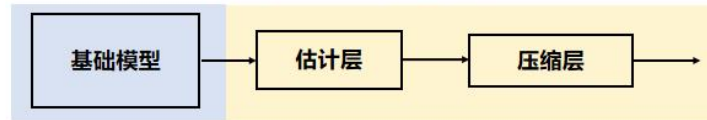
白皮书提出一种基于 AI 的准多模态信道信息获取基础模型设计，支持低开销信道估计、预测与压缩反馈这一系列典型用例。通过对信道估计、预测以及压缩反馈用例的研究，发现各用例所提取的信道特征存在相似部分，因此相应 AI 模型的设计具有轻量化的潜力，而更低的模型复杂度更加适配终端设备有限的存储与计算能力。

信道矩阵在空域、频域和时域上均具备丰富特征。在信道估计、预测与压缩反馈这三个应用中，模型都要对空频域信道特征进行提取，然而，三者与时域处理上有所不同：信道估计与压缩反馈针对已测量的时间段，其中信道估计可同时使用当前时刻前后信息，压缩反馈仅依赖历史及当前时刻信息；信道预测则面向未来时刻。基于这一差异，本白皮书提出一种新型 AI 模型架构，该架构包含两部分：第一部分为通用基础模型，用于提取空频域及部分时域共性特征。如图 6-17 所示，该模型以图像超分辨率网络 SRCNN 为基础，在特征提取、非线性映射和重建三阶段中引入含 LSTM 模块的残差结构，以增强时域特征提取能力，并缓解因模型加深而引发的梯度消失问题；第二部分为多任务适配层，主要针对不同任务提取空时频域的差异性特征。

该基础模型架构支持信道获取功能的灵活编排，如图 6-17 所示。由于基础 AI 模型已预先提取了各任务的共享特征，第二部分的多任务适配层在设计时可采用更少的参数量。同时，由于所有用例在推理时共享基础模型的输出结果，避免了推理过程中的重复计算，因此该架构能够显著降低 AI 模型的推理开销。实验表明，该设计可降低模型参数复杂度与推理计算复杂度达 50%以上。



(a) 准多模态AI模型应用于多任务



(b) 信道估计+信道压缩



(c) 信道估计+信道预测+信道压缩

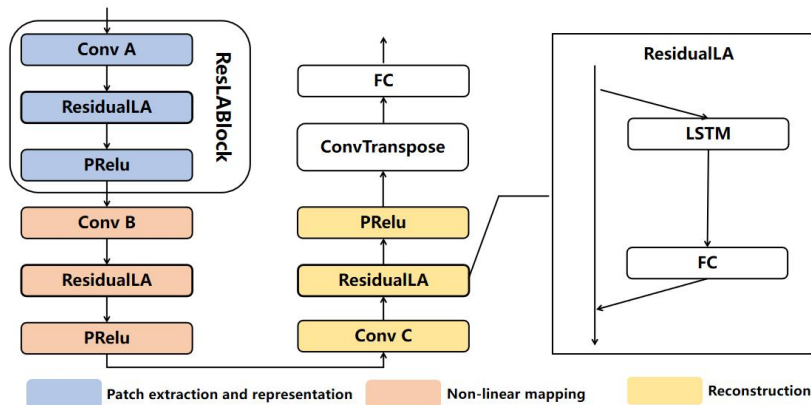


图 6-17 准多模态信道信息获取 AI 模型架构与基座模型设计

6.2. 用户调度与 MIMO 预编码

传统单小区和多小区用户调度与 MIMO 预编码技术在应对大规模用户和天线场景时，存在计算复杂度高、维度泛化能力不足等问题。作为 6G 实现全域谱效优化的核心环节，构建低复杂度、强泛化且支持跨区协同的智能调度与预编码机制，已成为释放多天线系统潜能的关键路径。本章将系统阐述基于 AI 的单小区和多小区用户调度与 MIMO 预编码技术，降低运算复杂度、提升小区谱效、降低小区边缘的干扰。

6.2.1. 单小区用户调度和干扰协调

6.2.1.1. 业界面临的挑战

用户调度与 MIMO 预编码是实现高谱效宽带多天线系统的关键。随着用户数、系统资源块数和天线规模的增加，可能的用户-资源块组合数量呈指数级增长，使得穷举搜索或贪婪算法的计算复杂度无法接受，采用启发式方法或分步优化可以降低复杂度，但往往导致系统资源分配效率低。

现有学习方法聚焦于窄带系统或单一策略学习，忽略了无线策略的置换特性，导致无法对时变的用户数进行泛化；也有方法启发式地采用强化学习，但只学习单一策略，且训练稳定性差、收敛速度慢。可以利用置换性质设计 GNN 使其具有维度泛化能力，但是现有 GNN 在处理相似用户信道时会出现“相同参数相同决策”的问题，即当两个用户信道条件相似时，GNN 会做出相同的调度决策；然而最优调度策略是非相同参数相同决策的，也就是只选择其中一个用户。如何解决 GNN 学习非相同参数相同决策尚未解决。

6.2.1.2. 思路与方法

针对宽带多用户 MIMO 系统的数模混合预编码，所有候选用户到基站的宽带信道为 $\mathbf{H} \in \mathbb{C}^{M \times K N_R \times N_T}$ ，调度矩阵为 $\mathbf{A} \in \{0, 1\}^{M \times K}$ ，模拟预编码和基带预编码分别为 $\mathbf{W}_{RF} \in \mathbb{C}^{N_T \times N_{RF}}$ 和 $\mathbf{W}_{BB} \in \mathbb{C}^{M \times N_{RF} \times K}$ ，用户的模拟组合器为 $\mathbf{v}_{RF} \in \mathbb{C}^{K N_R \times 1}$ 。联合优化需要学习从信道到所有优化变量的映射 $(\mathbf{A}, \mathbf{v}_{RF}, \mathbf{W}_{RF}, \mathbf{W}_{BB}) = f_J(\mathbf{H})$ ，但是由于未调度用户对应的基带预编码和模拟组合没有实际意义，这样直接学习导致神经网络不得不输出大量无效的变量。

- 联合学习结构设计

为了避免学习无效变量，提出了一种基于双 GNN 的联合学习框架，如下图 6-18 所示，基本思想是把复杂的联合优化问题分解为两个紧密耦合、相互嵌套的策略进行学习。内层是调度策略 $\mathbf{A} = f_S(\mathbf{H})$ ，外层是同时学习模拟预编码、对应调度用户的模拟组合器和基带预编码的策略 $(\mathbf{v}'_{RF}, \mathbf{W}'_{RF}, \mathbf{W}'_{BB}) = f_P(\mathbf{H}')$ ，其输入只包含调度用户的信道。两个策略由两个模块学习、通过一个不需要学习的、用于

从所有候选用的信道中提取调度用户的信道的可微分函数 $H' = f_H(H, A)$ 连接，确保梯度能够反向传播，从而实现端到端联合训练。

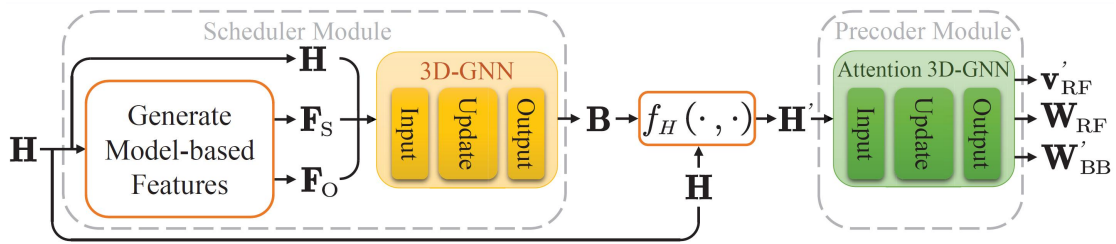


图 6-18 联合学习用户调度、混合预编码、模拟组合的结构图

这种设计既避免了学习无效决策（如对未调度用户的预编码），又保持了策略优化的整体性。整个架构支持无监督训练，可以直接以谱效（或能效等其他效用函数）最大化为目标函数，无需标签数据。

● GNN 设计

针对无线资源分配问题的结构，构建了两种类型的图结构：调度图和预编码图，如下图 6-19 所示。图中包含三类顶点：资源块顶点、基站天线顶点和用户天线顶点。超边连接这三个顶点类型，信道系数作为超边的特征。

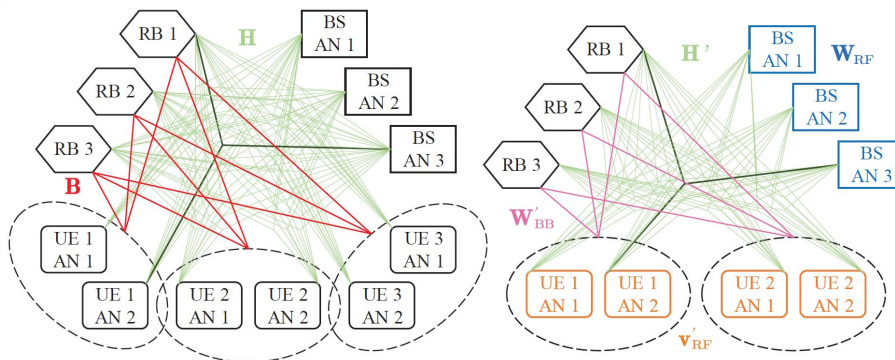


图 6-19 调度图和预编码图的建模

调度图中，动作定义为连接资源块顶点和用户天线顶点组的边，用基向量表示调度决策。在预编码图中，动作定义在不同类型的顶点和边上：模拟合并器为用户天线顶点的动作，模拟预编码器为基站天线顶点的动作，基带预编码器为资源块与用户天线组间边上的动作。这种基于图的表示方法可满足策略的置换性质。

● 调度模块设计

调度模块采用序列化 GNN 设计以解决相似用户信道的决策冲突问题。该模块由 K' 个连续的 3D-GNN 组成，每个 GNN 负责选择一个用户。这样可以保障后续 GNN 不仅接收原始信道信息，还接收前面 GNN 已做出的调度决策，从而确

保相似信道用户不会被重复选择。

每个 GNN 内部包含输入处理、更新过程和输出过程。输入过程引入了两种基于领域知识的特征：信道强度特征和正交性特征。信道强度特征捕获用户信道质量，正交性特征衡量用户间信道相关性，这两个特征增强了模型的泛化能力。更新过程采用线性聚合器，通过对相邻超边信息的加权聚合来更新隐藏表示。输出过程通过生成松弛的调度决策，在训练阶段保持可微性，在测试阶段转换为离散决策。这种设计有效解决了 GNN 与非相同参数相同决策之间的矛盾。

- MIMO 预编码模块设计

预编码模块采用带注意力机制的 GNN 来处理用户间的干扰问题。该模块包含注意力系数，可以衡量不同资源块上用户之间的干扰强度。在更新过程中，来自不同用户的信息通过注意力系数进行加权聚合，而其他类型的信息（资源块、天线等）则采用线性聚合。

输出过程通过压缩隐藏表示生成最终的预编码参数，并通过归一化操作满足功率约束和恒模约束。具体而言，基带预编码器通过全局功率归一化满足总功率约束，模拟预编码器和合并器通过元素级归一化满足恒模约束。

- 联合训练

白皮书设计了分阶段训练策略以解决嵌套策略学习的挑战。首先单独预训练预编码模块，用基于信道强度的简单调度生成训练样本。然后固定预编码模块，训练调度模块。最后进行联合微调，使两个模块协同优化。

- 推理复杂度

调度模块的计算复杂度为 $O(MKN_R N_T)$ ，预编码模块复杂度为 $O(MK'N_R N_T)$ ，均随天线数线性增长；而传统算法如半定松弛的复杂度为 $O((MKN_R N_T)^{3.5})$ ，可见所提出方法的推理复杂度远低于传统数值算法。

6.2.1.3. 成效与先进性

在 28GHz 频段 400MHz 带宽的典型城市宏基站场景下进行大量仿真评估，验证了所提出方法的性能和学习效率。

1) 高学习性能：所提方法的谱效非常接近当前最优的数值优化算法，并明显高于多种传统的次优算法。同时，其性能也远超其他基于深度学习的方法（如

图注意力网络（Graph Attention Network，GAT）、CNN 或前馈神经网络（Feedforward Neural Network，FNN））。

2) 可维度泛化：一个在特定系统配置（如 16 个资源块、30 个候选用户、16 根基站天线）下训练的模型，无需重新训练即可直接应用不同系统规模的场景，例如 RB 数在 4 到 128 之间、用户数在 3 到 60 之间、基站天线数在 8 到 128 之间动态变化的网络环境中，普遍能达到最优算法 90% 以上的性能。

3) 低复杂度：所提出方法的 CPU 推理时间与简单的非迭代算法相当，比复杂的数值算法快 50 至 100 倍；同时在所考虑的场景下所需的样本复杂度是 GAT 的 1%、CNN 的 0.4%，空间复杂度是 GAT 的 16%、CNN 的 7%。

6.2.2. 多小区用户调度和干扰协调

6.2.2.1. 业界面临的挑战

系统级多小区用户调度与干扰协调是一个复杂的多层优化问题，比如：1) 多用户的调度依赖于对信道状态的掌握，但在高负载下，系统难以实时掌握全局信道；2) 传统的多址接入方式在降低干扰和提升系统容量之间尚未寻找到最佳平衡点；3) 单小区调度策略面临着决策空间维度的增长远远超过用户数量增长的问题，多小区协作问题更严峻，传统方法难以很好适配 6G 的需求。

6.2.2.2. 思路和方法

针对上述难点，需要对现有主流方法，如 SRS 信道估计、多天线系统下的接入方式选择、快速调度算法及跨小区协作干扰消除等关键技术进行系统性的改善。AI 的引入为该目标的达成带来了新的实现途径。白皮书融合图神经网络、强化学习、知识图谱等 AI 技术与无线通信理论，从信道估计、多址技术、用户调度、干扰消除、多站点协作五大维度，提出系统性创新方案。

6.2.2.2.1. SRS 信道估计

在高用户负载场景下，传统密集 SRS 信道测量面临资源紧张与测量精度不足的挑战。为此，白皮书提出一种基于低密度信道测量并结合 AI 进行信道预测

与插值的新型信道获取方法，在降低导频开销的同时提升信道信息的完整性与准确性。该方法将信道的角度域信息表示为图像形式，并引入图像处理领域的深度学习去噪技术，充分挖掘信道角度域的稀疏特性，显著改善信道估计与插值的性能。其优势在于无需纯净信道数据作为训练标签，仅需在信道相干时间内采集同一信道受不同随机噪声影响的多个样本，即可完成网络训练。具体采用基于 N2N-RCAN 的信道插值算法，将导频子载波处的角度-频率域信道矩阵视作低分辨率图像，通过残差通道注意力网络执行超分辨率重建，以恢复全部子载波上的信道响应。针对实际系统中难以获取真实信道标签的难点，引入 Noise2Noise 学习机制进行无监督训练。为更好地提取信道稀疏性与深层特征，网络输入与输出均保持在角度-频率域中处理。

6.2.2.2.2. 多天线流间干扰消除和多址技术

方案中还融入一种新型的分层调制多址接入技术。受益于发射端的联合格雷调制以及相位对齐的预编码设计，所提方案将多用户数据拆分为一个共享核心数据流以及若干个拓展数据流，其中核心数据流以比特分复用的形式被多个具有流间干扰的用户所共享。在此基础之上，各用户还有一个专有的拓展数据流作为补充。尽管发射端发出符号的合成星座图是不规则的，但经信道传播后每个用户只会收到核心数据流及其自身所需的拓展数据流，且二者对应的合成星座图是一个规则的满足格雷映射的 H-QAM 星座图。此外，通过对发射端各比特流的欧氏距离参数进行优化，每个用户接收端的合成星座图具有不同的空间结构，这样各用户特定的规则星座图使得它们通过简单的信号处理流程实现较高速率成为可能。

6.2.2.2.3. 单站的快速多用户调度

调度过程中每个时隙之间是耦合的，前时隙的动作会影响后时隙的决策，构成马尔可夫过程，因为公平性函数的非线性，每个时隙的局部最优解无法构成全局最优解。优化李雅普诺夫虚拟队列稳定性及效用函数，保持每时隙李雅普诺夫最优，时间趋于无穷时长问题最优。

本方案利用李雅普诺夫优化框架，将长时调度问题解耦成单时隙调度问题。在单时隙调度中，采用基于剪枝图注意力神经网络的站内调度方法，将单小区中

AP 及其关联的用户建模为图中的节点。用户节点的特征包括：信道信息、用户权重、该用户与站内其他用户的信道相关情况。AP 与用户之间的边不包含特征，而用户之间的边则反映信道相关性。每个用户节点输出其被选中的概率，每个 AP 节点输出一个用户调度向量，向量中的一个元素对应于一个用户是否被调度。本方案的图可以根据信道相关性在每一层聚合和组合操作之前进行剪枝，避免过多邻居聚合特征所引发的过度平滑问题。

6.2.2.2.4. Cell-Free 网络用户干扰抑制方法

干扰协调模式在不同的场景和条件下各有优劣。白皮书提出两层闭环干扰协同框架：第 1 层由多智能体近端策略优化（Multi-Agent Proximal Policy Optimization, MAPPO）在部分可观测与时延失配条件下自适应选择干扰协调模式，面向系统吞吐与边缘公平性、回程与切换成本做多目标权衡；第 2 层在给定模式下，在多关系知识图谱上进行精细化功率分配，显式编码“服务/干扰/邻近/回程”等语义关系并捕捉长程干扰耦合。设计的多智能体强化学习算法，其中每个 UE 作为智能体处理相应的问题。这种协作多智能体任务通常可以视为一个部分可观察马尔可夫决策过程。每个 UE 首先将其观测输入到神经网络拟合的策略函数中，然后输出动作并生成奖励，完成与环境的交互。

白皮书提出一种新型功率分配框架：首先构建 Cell-Free 无线环境并生成数据，作为训练与评估基础。随后以知识图谱形式化建模网络，关系显式编码服务链路、干扰耦合、物理距离、AP-UE 连接等；基于关系感知注意力进行消息传递与聚合，捕捉长程依赖与全局耦合。可在下行链路上进行精细化、可扩展的功率分配优化，提升干扰抑制与系统性能。

6.2.2.2.5. 多站点的多用户协作调度

为实现高效的多站点协作调度，首先采用基于干扰感知的 AP 动态分簇机制。该机制基于 GNN 预调度结果，使每个 AP 通过预编码估算其对其他 AP 的潜在干扰，进而构建 AP 间干扰关系图，并采用基于干扰强度的聚类算法，将强干扰耦合的 AP 划分至同一协作簇。与传统方法相比，该机制可根据信道条件与用户分布自适应调整。当网络拓扑变化或用户移动时，分簇结果将动态更新，以确保

簇内 AP 始终保持最优的干扰协调效果。在完成基站分簇后，系统采用离散软演员评论家强化学习算法实现簇内多小区协同调度。每个基站作为一个智能体，同簇内多个智能体进行联合决策，以比例公平性为目标进行用户调度。智能体的状态空间包括：当前基站候选用户与簇内所有候选用户的信道相关性、用户自身信道增益、调度权重及历史调度信息。为降低复杂度，动作空间设计为对候选用户的离散选择或删除操作，相比连续动作空间显著提升了训练效率与收敛速度。该算法采用软策略梯度方法，有效平衡探索与利用。通过多智能体间的信息共享与协同学习，系统能够在复杂干扰环境中实现全局性能优化。

为实现高效的 RB 资源分配，提出一种以用户为中心的无蜂窝网络中接入点选择与资源块分配的联合优化方法。该方法将每个用户设备建模为自主智能体，基于本地观测的信道状态信息与历史平均速率，利用多智能体近端策略优化框架进行训练和决策：智能体输出包括二进制 AP 选择向量与离散 RB 索引的复合动作，通过集中式 CPU 仲裁机制解决资源竞争，目标是用户和速率最大化。在训练阶段，采用参数共享架构，所有智能体共用策略网络与价值网络，通过集中式训练学习协作策略，有效缓解重叠簇引入的站间干扰；执行阶段则仅依赖本地状态进行分散决策，避免通信开销。仿真结果表明，所提方法在系统和谱效上显著优于传统方案——相较于纯随机策略、LLSF（Largest Large Scale Fading，最大大尺度衰落）+RR（Round-Robin，轮询）与 LLSF+PF（Proportional Fair，比例公平），其用户和速率分别提升 243%、96%与 39%，并在不同发射功率下均保持优越的干扰协调能力与网络吞吐量。

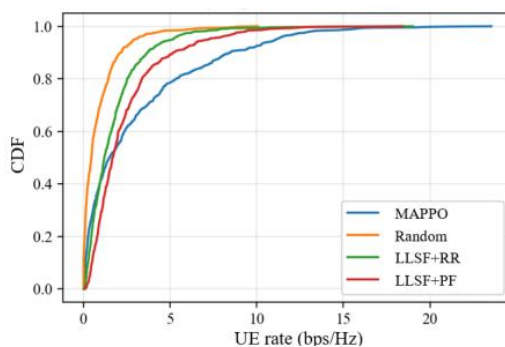


图 6-20 所提方案与基线方案用户速率的累积分布函数（Cumulative Distribution Function, CDF）曲线

6.2.2.3. 成效与先进性

6.2.2.3.1. SRS 信道估计

1) 提升信道估计性能。针对大规模 MIMO 场景下低密度的信道测量结果，可降低平均信道估计误差，尤其在低信噪比区间，可使信道估计性能提升 3dB。

2) 降低导频开销。信道估计在高维信道情况下计算复杂度高，开销较大，本方案减少导频长度带来开销减少，可降低 50%。

3) 缩短训练时间。引入去噪神经网络进行信道角度域特征提取，无需获取纯净的信道数据，对数据进行处理后可缩短训练时间，计算时间降至原来的 1.5%。

6.2.2.3.2. 单站快速调度和新型分层调制多址技术应用

采用 GNN 图神经网络对站内用户进行预调度，新型分层调制多址接入和优化的预编码方法，仿真如下图 6-21 和图 6-22，20%分位用户的长期平均速率相比于基线方案提升 2.14 倍，20%分位的用户信干噪比相比于基线方案提升至少 5.26dB。

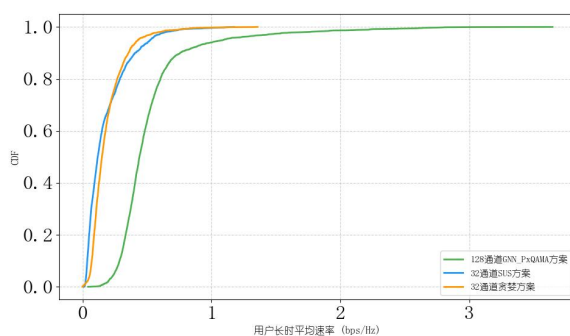


图 6-21 所提方案与基线方案的长期平均用户速率的 CDF 曲线

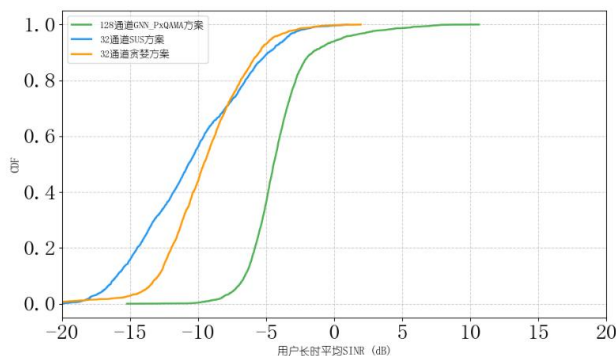


图 6-22 所提方案与基线方案的长期平均信干噪比的 CDF 曲线

6.2.2.3.3. Cell-Free 网络用户干扰抑制方法

基于 MAPPO 算法的用户干扰水平显著降低，尤其对于高干扰的用户来说有明显改善。从箱线图中可以看出对于干扰 20%分位的用户，MAPPO 联合决策对比阈值法+动态成簇大约降低干扰 2dB；对于干扰 50%分位的用户，MAPPO 联合决策对比阈值法+动态成簇大约降低干扰 1.5dB。由此可知，MAPPO 的联合决策方案对干扰强的用户提升更明显。

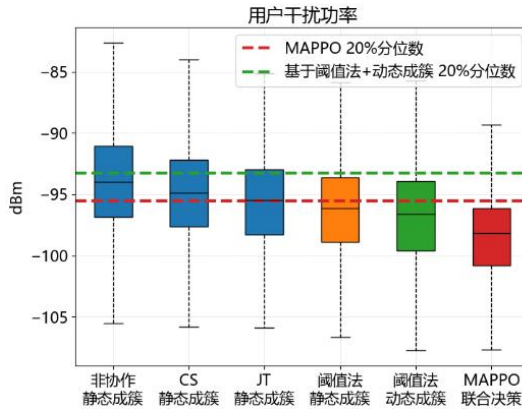


图 6-23 用户所受干扰功率的箱线图

6.2.2.3.4. 多站点的多用户协作

本方法首先利用 GNN 预调度的结果，让每个 AP 执行预编码操作以估算对其他 AP 产生的潜在干扰。基于干扰估计结果，构建 AP 间的干扰关系图，并采用基于干扰强度的聚类算法将干扰耦合较强的 AP 划分到同一协作簇中。在完成基站分簇后，采用离散软演员评论家强化学习算法实现簇内的多小区协同调度。下图仿真显示，20%分位用户的长期平均速率在所提方案下相比于基线方案提升 2.1 倍，20%分位用户信干噪比提升 6.24dB。

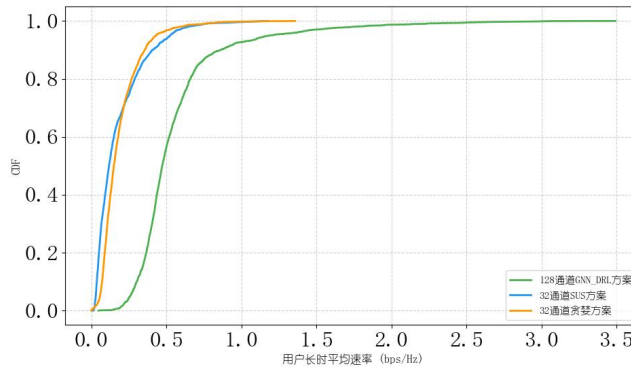


图 6-24 所提方案与基线方案的长期平均用户速率的 CDF 曲线

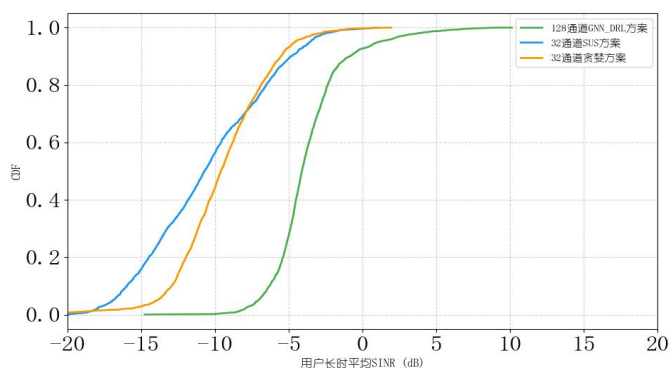


图 6-25 所提方案与基线方案的长期平均信干噪比的 CDF 曲线

该方法不仅在理论上实现了重要突破，更为 6G 网络的工程实现提供了可行的技术路径。通过仿真平台验证，方案的实用性和有效性得到了证明。

6.3. 空口 AI 编译码和波形设计方法

传统信源与信道编码通常基于信息等重要性假设，在任务导向的通信场景中无法适应多模态信源的非均匀特性，产生显著冗余；OFDM 波形的循环前缀 CP 属于冗余开销，现有无 CP 方案在性能、复杂度和灵活性方面仍有明显不足。本章将系统阐述基于无速率编解码与 AI 辅助波形设计的创新方法，通过引入语义特征提取与自适应均衡技术，实现系统谱效与鲁棒性的提升。

6.3.1. 信源信道联合设计方法

6.3.1.1. 业界面临的挑战

在新型业务的驱动下，网络面临超高吞吐、超低时延与极高可靠性的严峻挑战。传统信源与信道编码通常基于信息等重要性假设，在任务导向场景中无法适应多模态信源的非均匀特性，产生显著冗余。为提升谱效，挖掘语义层次的关键信息可实现高效通信。深度学习驱动的信源信道联合编码就是可实现任务导向的高效通信。其利用 AI 进行信源特征提取与信道自适应编码，直接映射至语义空间，显著提升传输效率与鲁棒性。现有研究在图像重建及动态信道适应等方面已展现优越性能，但面对未知或时变信道时仍缺乏泛化与自适应能力。

6.3.1.2. 思路和方法

白皮书设计了一种具有无速率特性的信源信道联合编解码方案，将其应用于图像重建任务，以在保持信道自适应特性的同时显著提升通信效率，该方案的核心思想是基于多项式-高斯分布构建了一种新型变分自编码器（Variational Autoencoder，VAE），称为多项式-高斯变分自编码器（Multinomial Gaussian-Variational Autoencoder, MG-VAE）。通过为深度联合信源信道编码（Joint Source-Channel Coding，JSCC）框架引入无速率特性，使系统能够自适应未知或动态网络条件。白皮书提出的基于 MG-VAE 的无速率图像信源信道联合编码系统的编码器与解码器示意图如下图 6-26 所示。

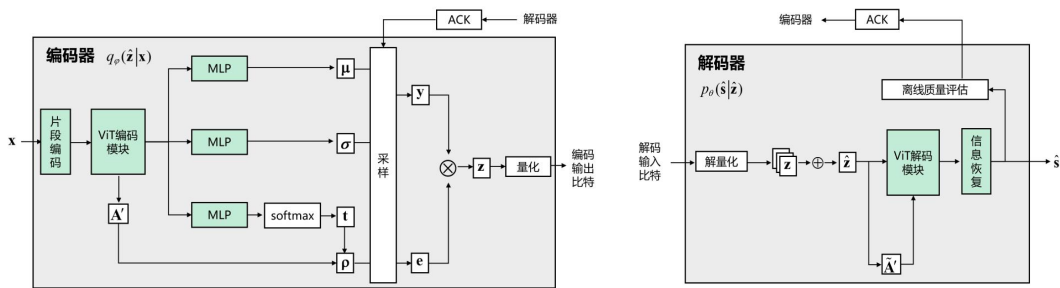


图 6-26 基于 MG-VAE 的无速率信源信道联合编码器与解码器

6.3.1.3. 成效和先进性

在完备知识库背景下各方案峰值信噪比（Peak Signal-to-Noise Ratio, PSNR）性能对比曲线如下图 6-27 所示。在 SNR 为-1~27dB 范围内，所提方案的 PSNR 不低于 30dB，对应高质量图像重建标准。

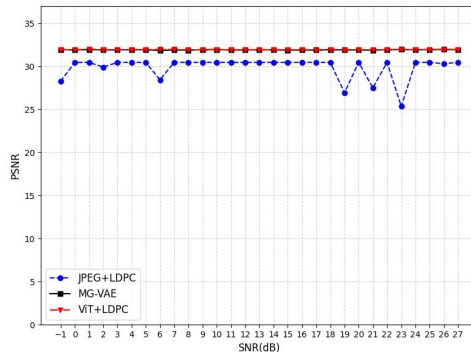


图 6-27 图像重建性能 PSNR 对比图

各方案较 R16 方案的等效信道容量增益倍数（相对值）曲线如下图 6-28 所示。

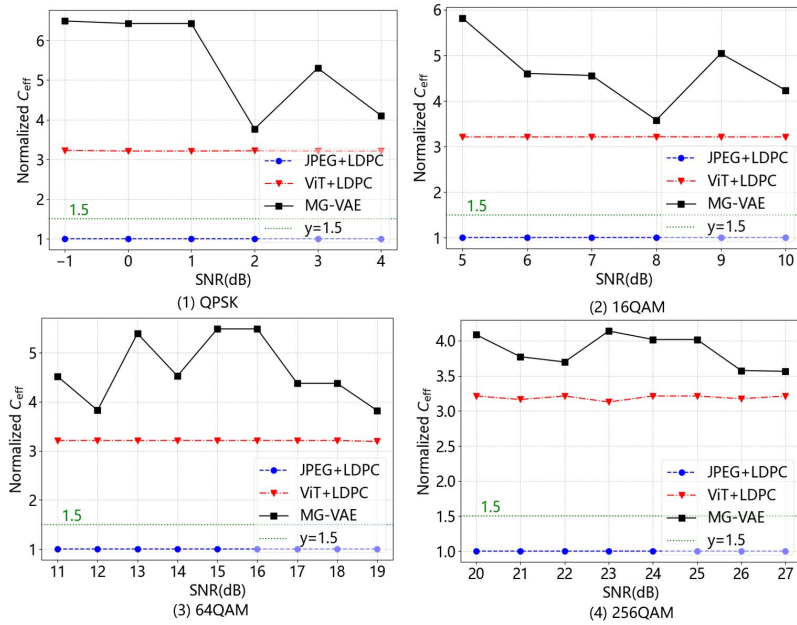


图 6-28 等效信道容量增益倍数对比图

在 PSNR 不低于 30dB 的情况下，所提完备知识库方案可在 SNR 介于 -1dB~27dB 时较 R16 标准的 JPEG+LDPC 方案具有 3.5~6.4 倍信道容量增益。

6.3.2. 无 CP OFDM 波形设计

6.3.2.1. 业界面临的挑战

OFDM 系统中由于多径传播引起的符号间干扰（Inter-Symbol Interference, ISI）和载波间干扰（Inter-Carrier Interference, ICI）通常通过插入 CP 来缓解，但 CP 属于冗余开销，在高速率场景下损耗可达 25%，成为瓶颈。现有无 CP 方案如均衡器设计和广义频分复用（Generalized Frequency Division Multiplexing, GFDM）在性能、复杂度和灵活性方面仍有明显不足。近年来的深度学习方法虽展现潜力，但在无 CP 环境下鲁棒性和实时性欠佳，且多集中于接收端，缺乏收发机一体化设计，难以应对复杂时频双选信道的需求。

6.3.2.2. 思路与方法

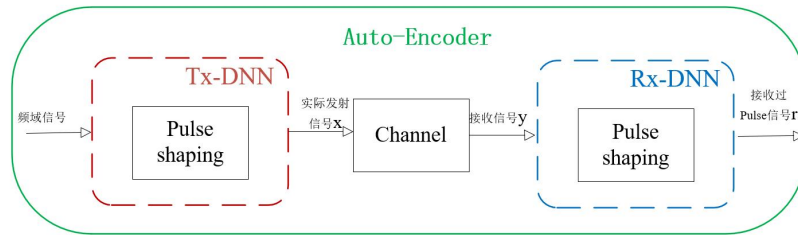


图 6-29 AI 辅助的 PS-OFDM 原理图

白皮书提出了 AI 辅助的脉冲成形 OFDM，通过收发端联合设计来摆脱对 CP 的依赖。在发射端，利用深度学习实现脉冲成形滤波，改善信号的时频局域性，以减轻 ISI 和 ICI；在接收端，引入卷积神经网络均衡解调网络，有效捕捉时变信道的相关性，提升解调精度并降低导频开销。此外，系统结合离线训练与在线微调，实现对信道动态适应，在不同传输场景下保持鲁棒性和高效性。

6.3.2.3. 成效和先进性

仿真结果如下图 6-30 所示，在时频双选信道下，即使在 50–100 km/h 高速移动场景中，系统对多普勒频偏几乎不敏感，依旧保持稳定性能，谱效相比传统方案提升约 15.5%。所提方法不仅突破了 CP 机制的限制，还充分利用了 AI 的自适应能力，在提升谱效的同时显著增强了系统在复杂信道环境下的鲁棒性，展现出在高速移动场景中的应用潜力。

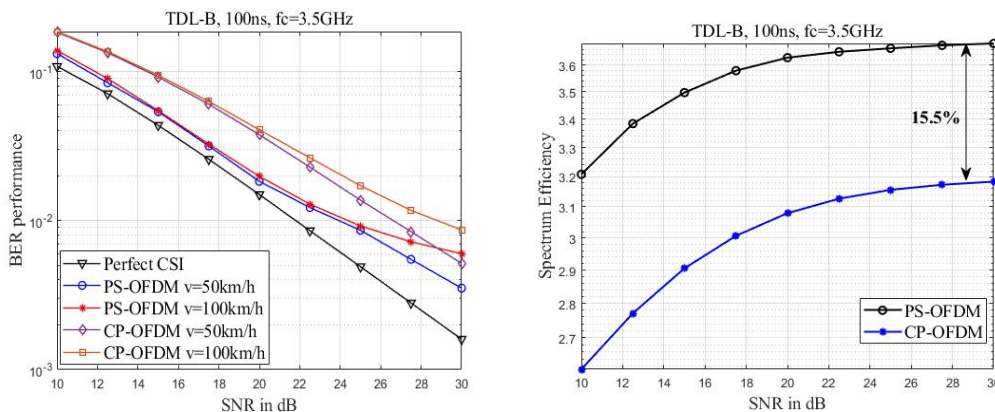


图 6-30 时频双选信道下误码率与谱效性能仿真图

6.4. 空口 AI 多模块联合设计方法

当前，空口 AI 模型设计正面临模型冗余、资源开销高昂与协同效率低下等核心挑战。业界普遍采用针对单一功能独立设计模型的路径，导致训练、存储及推理成本叠加，且模块间割裂优化限制了系统性能的进一步提升，构建高效、轻量与可泛化的 AI 模型架构已成为技术演进与规模部署的必然要求。本章将系统介绍发射机多模块、接收机多模块联合优化的典型场景中，如何通过多模块化联合设计的创新方法，实现模型复杂度的显著降低与系统性能的协同跃升。

6.4.1. 发射机多模块联合设计

6.4.1.1. 业界面临的挑战

在大规模多用户 MIMO 系统中，预编码等高维发射模块的计算复杂度和为了获得准确信道所需的导频开销都随着天线和用户数快速增长，导致难以实时实现数值算法，且随着信道的过时谱效急剧下降。另外，传统发射机对信道估计、信道预测与混合预编码等多个模块进行独立优化，这种分而治之设计模式中各模块优化目标不一致、使系统总体性能下降或导频开销增加。

现有联合优化多模块的学习方法没有建模端到端优化问题，且往往给定部分模块优化其他模块，并没有实现真正的联合优化。其次，现有方法通常采用的深度神经网络（Deep Neural Network, DNN）不能维度泛化，无法适配用户和子载波数的动态变化。最后，现有方法没有考虑 TDD 系统中上下行子帧比例动态调整的需求。

6.4.1.2. 思路与方法

白皮书建立了一种主动优化学习框架，无需显式信道估计与信道预测模块，建模直接基于上行导频信号学习“多下行子帧最优预编码器”的优化问题，通过学习从上行 SRS 导频信号到下行多个子帧预编码器的映射关系，实现“最大化所有下行子帧期望总速率”的目标。以模拟与数字混合预编码为例，基于上行导频优化所有下行子帧混合预编码矩阵的优化目标可建模为：

$$\max_{\bar{\mathbf{W}}_{RF}, \bar{\mathbf{W}}_{BB}} E_{\bar{\mathbf{H}}|\mathbf{R}} \left\{ \sum_{l=1}^{L_{DL}} U(\mathbf{H}^l, \mathbf{W}_{RF}^l, \mathbf{W}_{BB}^l) \right\}$$

其中 $U(\cdot)$ 为单帧和数据率函数， $\mathbf{H}^l \in \mathbb{C}^{N_t \times N_{RF}}$ ， $\mathbf{W}_{RF}^l \in \mathbb{C}^{N_t \times N_{RF}}$ ， $\mathbf{W}_{BB}^l \in \mathbb{C}^{N_{RF} \times K}$ 表示第 l 帧的信道、模拟和数字预编码矩阵， $\bar{\mathbf{H}} = [\mathbf{H}^1; \dots; \mathbf{H}^{L_{DL}}]$ ， $\bar{\mathbf{W}}_{RF} = \{\mathbf{W}_{RF}^1, \dots, \mathbf{W}_{RF}^{L_{DL}}\}$ 和 $\bar{\mathbf{W}}_{BB} = \{\mathbf{W}_{BB}^1, \dots, \mathbf{W}_{BB}^{L_{DL}}\}$ 表示多子帧信道、模拟和数字预编码矩阵张量， L_{DL} 为下行子帧个数， \mathbf{R} 为上行导频矩阵， $E_{\bar{\mathbf{H}}|\mathbf{R}}$ 表示基于 \mathbf{R} 的信道条件期望。

直接求解存在两个挑战：条件期望 $E_{\bar{\mathbf{H}}|\mathbf{R}}$ 依赖未知的信道条件概率密度函数 $f(\bar{\mathbf{H}}|\mathbf{R})$ ，无法直接求解；多子帧变量 $\bar{\mathbf{W}}_{RF}$ 、 $\bar{\mathbf{W}}_{BB}$ 相互耦合，求解复杂度随子帧数 L_{DL} 指数增长。为了解决第一个问题，主动优化通过利用期望的全概率公式把原问题的参数优化问题转化为泛函优化问题，优化目标变为：

$$\max_{\bar{\mathbf{W}}(\mathbf{R})} E_{\bar{\mathbf{H}}|\mathbf{R}} \left\{ \sum_{l=1}^{L_{DL}} U(\mathbf{H}^l, \mathbf{W}^l(\mathbf{R})) \right\}$$

其中 $\mathbf{W}^l(\mathbf{R})$ 表示从 \mathbf{R} 映射到第 l 子帧预编码器的策略。

此时，无需获取未知的信道条件概率密度函数，只需生成 $\{\bar{\mathbf{H}}, \mathbf{R}\}$ 样本，即可通过无模型无监督学习方法得到混合预编码策略。

为了解决第二个问题，进一步把泛函优化问题分解为 L_{DL} 个独立的“单子帧速率最大化”：每个子问题对应单个下行子帧，针对每个子问题，设计专用 GNN 学习该子帧的预编码映射，最终构成并行主动优化网络，其结构如下图 6-31 所示。

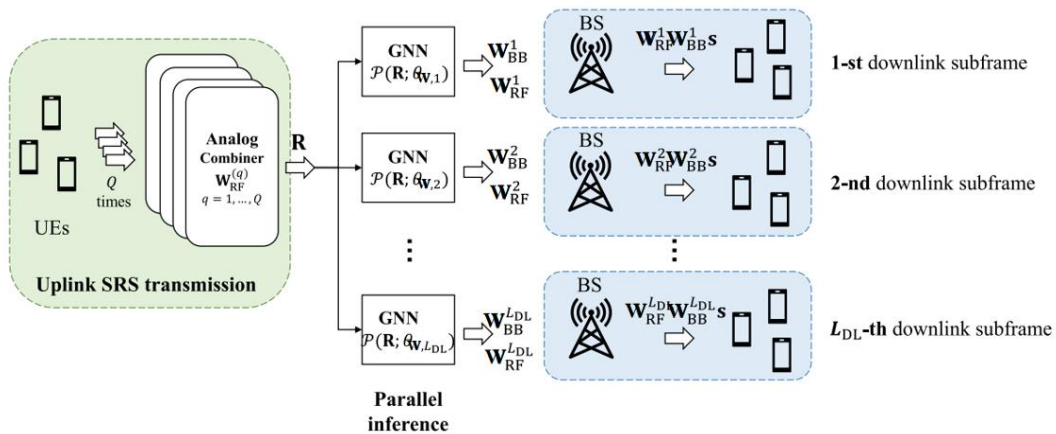


图 6-31 并行主动优化 DNN 结构

该架构为每一个下行子帧配备一个独立的有注意力机制的 GNN 模块，所有模块共享同一个上行 SRS 输入。当子帧数变化时，只需增加或减少 GNN 模块数即可。GNN 利用无线策略的置换性质获得了用户和子载波维度的泛化能力。

6.4.1.3. 成效与先进性

1) 高学习性能：在导频长度为 9 时，所提出 DNN 与传统分模块设计（匹配追踪用于信道估计、深度学习用于信道预测、相位提取用于混合预编码优化）相比，在不同信噪比下均达到更高的和数据率，并且在低信噪比场景下性能优势更加显著；随导频长度增加，所提出 DNN 的学习性能接近理想 CSI 下的数值算法。在子帧上性能良好，且在用户速度（30km/h 至 300km/h）不同时性能良好。

2) 可维度泛化：能对用户数泛化，采用 $K=3$ 样本训练的模型，无需微调即可在用户数 $K=2\sim 6$ 时直接推理、且性能损失小于 2%；能对子载波数泛化，采用 $M=4$ 子载波的样本训练的宽带模型，可直接在子载波为 $M=2\sim 128$ 时进行推理。

3) 低导频开销与复杂度：为达到理想 CSI 下的数值算法 95% 的性能，所需的导频长度比传统方法减少 50% 以上。窄带系统的 CPU 推理时间为 29.2ms，宽带系统的 CPU 推理时间为 44.4ms；经过 GPU 加速后，推理时间降至 7.7ms。模型参数量随系统规模增长，而传统学习方法的空间复杂度随子载波数线性增长。

6.4.2. 接收机多模块联合设计

6.4.2.1. 业界面临的挑战

在无线通信系统中，发送信号会受到发射器件的非线性特性、时变频率选择性衰落信道、外界干扰以及器件热噪声等多重因素的影响，导致波形畸变，进而严重影响信道估计、均衡与解调等环节的性能。传统接收机通常对信道估计、均衡和解调等模块分别进行独立优化，这种分段处理方式往往难以实现全局最优，导致整体解调性能受限。近年来，业界开始探索使用 AI 模型替代传统基于 DMRS（解调参考信号）的信道估计、插值、均衡与解调等模块。该类方法充分发挥 AI 模型的非线性拟合与去噪能力，不仅能够更准确地进行信道估计，还能有效补偿由非线性器件引入的信号畸变。同时，AI 模型可借助符号级非导频数据实

现一定程度的信道盲估计，从而进一步提升信道估计的稳健性，降低数据解调误码率，提高系统谱效。

然而，现有 DMRS 导频设计仍存在开销过大的问题，尤其在利用 AI 接收机对 DMRS 导频样式进行联合优化方面，目前业界尚缺乏有效解决方案。在中高速移动场景下，由于信道变化剧烈，通常需在时域配置多列 DMRS 导频序列以保障估计精度，但这会导致导频在时频资源中占比过高，反而限制谱效的提升。现有 5G 双列 DMRS 导频结构最多仅支持 24 流，若 6G 系统的流数进一步增加，则需成倍扩充时频资源用于导频传输，这将显著提高非数据资源的比例，造成谱效的严重下降。

6.4.2.2. 思路、方法与成效

面向多模块 AI 接收机的设计，业界通常采用单一 AI 模型替代由 DMRS 信道估计、均衡及解调等多个传统模块的方案。该 AI 接收机模型可选用 CNN、残差网络（Residual Network, ResNet）或 Transformer 等主流结构。其输入通常包含发送端导频（正交或非正交）、接收端基于正交导频的信道估计信息以及接收到的时频域符号级数据；输出则为比特流中每位为 0 或 1 的概率值。模型训练需以发送端的原始比特流作为标签数据进行监督学习。

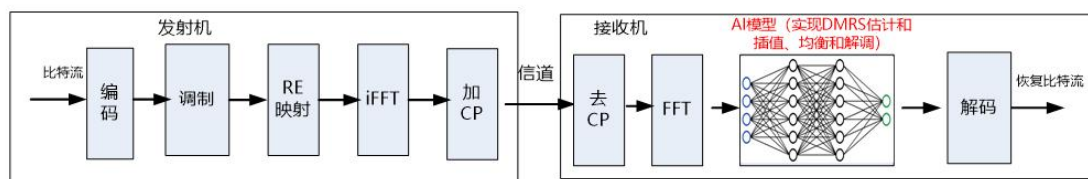


图 6-32 基于 AI 的接收机流程

针对中高速或多流场景下导频开销过高、规则导频分布导致信道信息冗余的问题，白皮书提出采用 AI 技术自动优化非规则稀疏导频图案的设计方案。该方案能依据信道在时域和频域的动态特性，自适应地寻优非规则导频分布样式。例如，在低移动性的非视距场景中，多普勒频偏影响较小，但频域多径衰落显著，因此可在时域配置稀疏的非规则正交导频，而在频域配置相对密集的非规则正交导频。对于中高速场景，由于多普勒效应影响更大，则需在时域上配置更密集的非规则正交导频。为进一步增强解调性能，还可引入非正交导频；通常距离正交

导频越近的位置，非正交导频的功率配比可设置得越低。基站可采用参考信号位图来指示非规则导频的位置：例如位图中的“0”表示该资源单元传输数据，“1”表示传输正交导频，这种方式会引入一定的空口信令开销。

为降低导频空口信令配置开销，可根据不同场景的信道变化特征设计半静态导频样式。在此设计中，正交导频在时频域中保持完全正交与稀疏分布，并采用固定排列形式（如阶梯状或对角状），使重复出现的导频序列在时频资源上实现完全错开。这种分布方式的优势在于，以较低的开销使正交导频在频域和时域上出现次数最大化，从而更充分地捕捉信道在频率上的衰落特性与时域上的变化特征。该设计在同时保障频域和时域采样密度前提下实现了更稀疏的导频分布，但也增加了传统信道估计算法的难度。借助多模块 AI 接收机的非线性拟合能力，可在提升信道估计精度的同时进一步降低导频开销。

经仿真验证，该非规则稀疏导频样式可节省约 50% 的导频开销，并借助 AI 接收机提升数据解调性能，使误比特率降低 2~3 dB 以上。在移动速度较快、信道波动较大的场景下，该方案可实现 30% 以上的谱效提升。

7. 空口 AI 多模块集成与系统级仿真

本章将介绍空口 AI 多模块的集成方法，以实现系统谱效、能效、复杂度、小区间干扰消除等多维度指标，并通过系统级仿真进行综合验证。

7.1. 业界面临的挑战

当前，空口 AI 技术的研究与应用仍主要聚焦于单一通信模块的性能优化，尚未形成系统化的端到端设计能力。由于空口协议栈各模块间耦合紧密、接口复杂，构建完整 AI 空口系统面临极高的工程复杂度和算法设计挑战，导致相关方案在现有平台上难以实用化部署。与此同时，空口性能的评价维度多元，既包含容量、谱效、能效等传统指标，也涉及算法复杂度、边缘用户干扰抑制等与 AI 能力密切关联的新型维度。在此背景下，如何从众多 AI 增强模块中筛选出关键组成部分，并将其有机融入现有传输链路，进而通过高可信度的系统级仿真对多维性能指标进行协同验证与优化，已成为推动智能空口技术从理论走向规模化部

署的核心挑战。

7.2. 思路和方法

本白皮书设计了一套空口 AI 集成方案，该方案将单模块性能优化后获得的链路-系统平台接口曲线以及多个 AI 集成模块整合至统一的系统级仿真平台中，最终通过多模块协同的系统级仿真，全面验证多个维度的空口性能表现。

下图 7-1 展示了空口 AI 集成方案中各模块的调用关系。所需集成的链路-系统平台接口曲线包括三个外部模块和两个集成模块。三个外部模块分别为：基于速度与导频密度的发射端信道误差模型、基于速度与导频密度的接收端信道 CSI 误差模型，以及不同信噪比（SNR）条件下的器件非线性补偿增益映射表；两个集成模块分别为：单小区调度与 MIMO 预编码集成模块，以及多小区联合调度与预编码集成模块。

在系统级仿真平台的终端模块中，通过上行 sounding 功能引入发射端信道误差模型。相较于非 AI 技术，基站侧采用基于 AI 的信道预测能够获得更精准的信道信息，从而有效提升 MIMO 发射预编码矩阵的精度，并进一步优化下行吞吐量。接收端信道 CSI 误差模型则反映了基于 AI 的 DMRS 估计对接收端 CSI 精度的影响，进而作用于接收机整体性能。不同 SNR 下的非线性补偿增益映射表，体现了基于 AI 的非线性补偿为接收端信干噪比带来的增益。上述三个 AI 模块对提升系统容量、谱效及能效等关键指标具有重要作用。

单小区调度与 MIMO 预编码集成模块嵌入基站侧的单小区调度功能中，突显了基于 AI 的调度与预编码在降低算法复杂度方面的优势。多小区联合调度与预编码集成模块则集成于基站的多小区联合调度功能中，展现了基于 AI 的联合调度与预编码在抑制小区间干扰、改善边缘用户性能方面的作用。

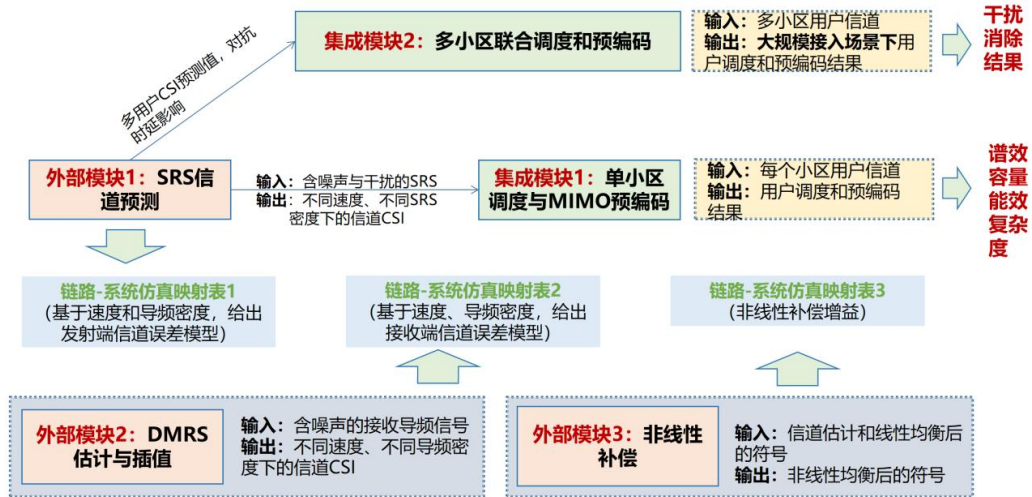


图 7-1 空口 AI 集成方案模块调用关系图

7.3. 成效和先进性

本小节主要介绍空口 AI 的具体仿真配置，仿真数据结果及相应的分析。仿真主要是完成不同 UE 移动速度配置下，对 5G non AI、6G non AI 以及 6G AI 三种集成方案的性能进行对比验证。三种集成方案的具体配置如下：

- Case 1: 5G non AI
 - ✓ BS 天线配置为 32 TXRU, $(M,N,P,Mg,Ng,Mp,Np) = (2,8,2,1,1,2,8)$;
 - ✓ UE 天线配置为 2 TXRU, $(M,N,P,Mg,Ng,Mp,Np) = (1,1,2,1,1,1,1)$;
 - ✓ 引入传统的基于速度和导频密度的发射端信道误差模型，和基于速度、导频密度的接收端信道 CSI 误差模型。
- Case 2: 6G non AI
 - ✓ BS 天线配置为 128 TXRU, $(M,N,P,Mg,Ng,Mp,Np) = (12,16,2,1,1,4,16)$;
 - ✓ UE 天线配置为 4 TXRU, $(M,N,P,Mg,Ng,Mp,Np) = (1,2,2,1,1,1,2)$;
 - ✓ 引入传统的基于速度和导频密度的发射端信道误差模型，和基于速度、导频密度的接收端信道 CSI 误差模型。
- Case 3: 6G AI
 - ✓ BS 天线配置为 128 TXRU, $(M,N,P,Mg,Ng,Mp,Np) = (12,16,2,1,1,4,16)$;
 - ✓ UE 天线配置为 4 TXRU, $(M,N,P,Mg,Ng,Mp,Np) = (1,2,2,1,1,1,2)$;
 - ✓ 引入使用 AI 模型进行调优的速度和导频密度的发射端信道误差模型，

和基于速度、导频密度的接收端信道 CSI 误差模型，同时引入不同 SNR 下的非线性补偿增益映射表，并集成单小区调度和 MIMO 预编码集成模块和多小区联合调度和预编码集成模块。

仿真数据结果如表 7-1 和图 7-2 所示，仿真数据结果包括 5 种 UE 移动速度，分别为 3km/h、30km/h、60km/h、120km/h、350km/h。从仿真数据结果可以看出，在低速情况下，6G non AI 的 Case2 相比于 5G non AI 的 Case1，具有明显的优势，然而在中高速情况下，Case2 的优势明显减弱；此时，对于引入基于 AI 调优的信道误差模型和 MIMO 预编码集成模块的 6G AI Case3，在中高速情况下仍然能够保持明显的优势，充分展现了 AI 模型的自适应能力以及在中高速移动场景中的应用潜力。

表 7-1 空口 AI 仿真数据结果表

UE 移动速度	UE 天线	BS 天线	方案配置	小区平均谱效 (bps/Hz)
3km/h	2TXRU	32TXRU_32ant	Case1: 5G non AI	8.82
	4TXRU	128TXRU_384ant	Case2: 6G non AI	24.59
			Case3: 6G AI	25.31
30km/h	2TXRU	32TXRU_32ant	Case1: 5G non AI	3.59
	4TXRU	128TXRU_384ant	Case2: 6G non AI	8.05
			Case3: 6G AI	20.58
60km/h	2TXRU	32TXRU_32ant	Case1: 5G non AI	3.51
	4TXRU	128TXRU_384ant	Case2: 6G non AI	6.23
			Case3: 6G AI	19.58
120km/h	2TXRU	32TXRU_32ant	Case1: 5G non AI	3.66
	4TXRU	128TXRU_384ant	Case2: 6G non AI	6.93
			Case3: 6G AI	19.72
360km/h	2TXRU	32TXRU_32ant	Case1: 5G non AI	3.30
	4TXRU	128TXRU_384ant	Case2: 6G non AI	11.91
			Case3: 6G AI	19.40

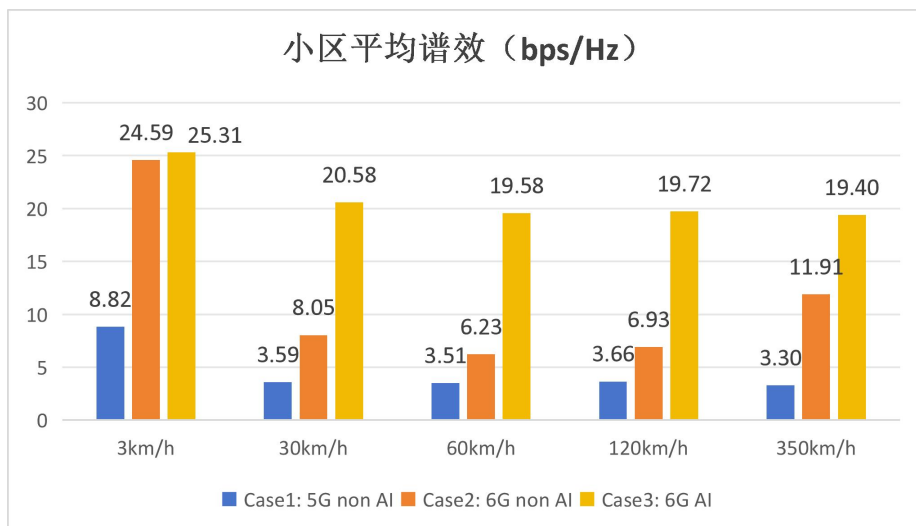


图 7-2 空口 AI 仿真数据结果柱状图

基于系统仿真平台得到的谱效结果，结合基站的功耗模型可以进一步得到系统的能效。此外，还统计了单小区调度和 MIMO 预编码 AI 算法以及基线算法的复杂度、多小区联合调度和 MIMO 预编码 AI 算法以及多小区联合调度基线算法的复杂度。相比 5G,6G AI 系统以 5 倍的带宽和 4 倍的收发通道数，实现了 2 倍谱效、10 倍能效、10 倍容量、算法复杂度的增加不超过 5 倍，多 AI 模块的集成方案实现了系统性能的综合提升。

8. 空口 AI 样机

为了在实际场景中验证前述关键技术，项目组搭建了集中式、分布式大规模 MIMO 样机和超高阶调制样机，本章将对三套样机进行简要介绍。

8.1. 集中式大规模 MIMO 样机

集中式超大规模 MIMO 样机采用 6GHz 频段对 6G 空口 AI 技术展开验证，样机系统架构如下图 8-1 所示。

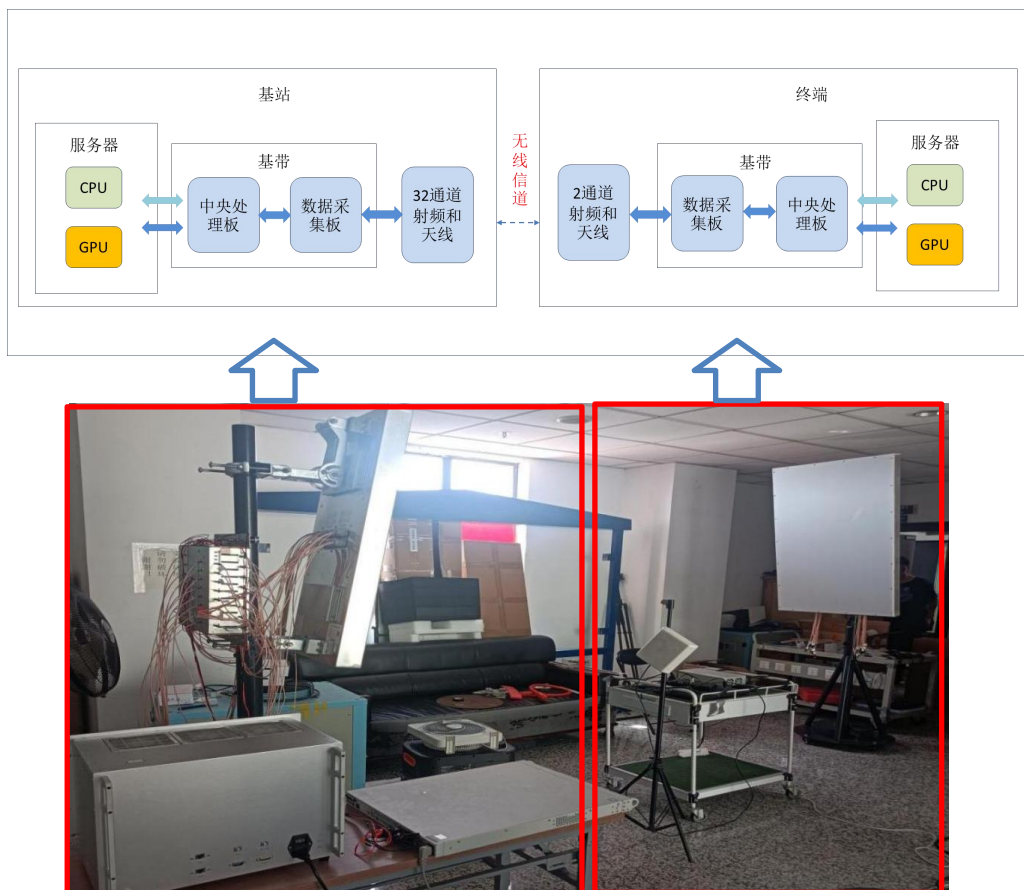


图 8-1 集中式超大规模 MIMO 样机系统架构与实物图

样机系统包括基站端和用户端两部分，二者均由服务器、基带、射频前端和天线阵列组成，各模块配置和功能如下：

1) 集中式射频和天线单元：基站侧射频和天线单元支持 32 独立收发通道，终端侧射频和天线单元支持 2 独立收发通道，负责完成射频信号到空口电磁波信号转换。射频和天线单元通过射频线连接到基带单元。

2) 基带：基带主要包含中央处理板和数据采集板两部分，其中数据采集板负责训练、推理等数据的收集工作，中央处理版负责多信道数据汇聚与信道的初步估计。基带通过网口和光口连接到服务器。

3) 服务器：服务器主要包含 GPU、CPU 两部分。其中 GPU 可依据基带采集的训练数据进行 AI 模型的训练工作。CPU 可完成模型推理、模型管理功能。

集中式超大规模 MIMO 样机系统可支持 32×2 组独立的上下行信道，具有信道相关数据的收集、处理能力，能够满足 6G 空口 AI 技术验证的高空间复杂度、高速率、高业务容量、高算力的要求。通过此样机完成了对基于时频空相关性的 AI 信道压缩反馈方案以及基于上下行信道部分互易性增强方案的性能验

证。

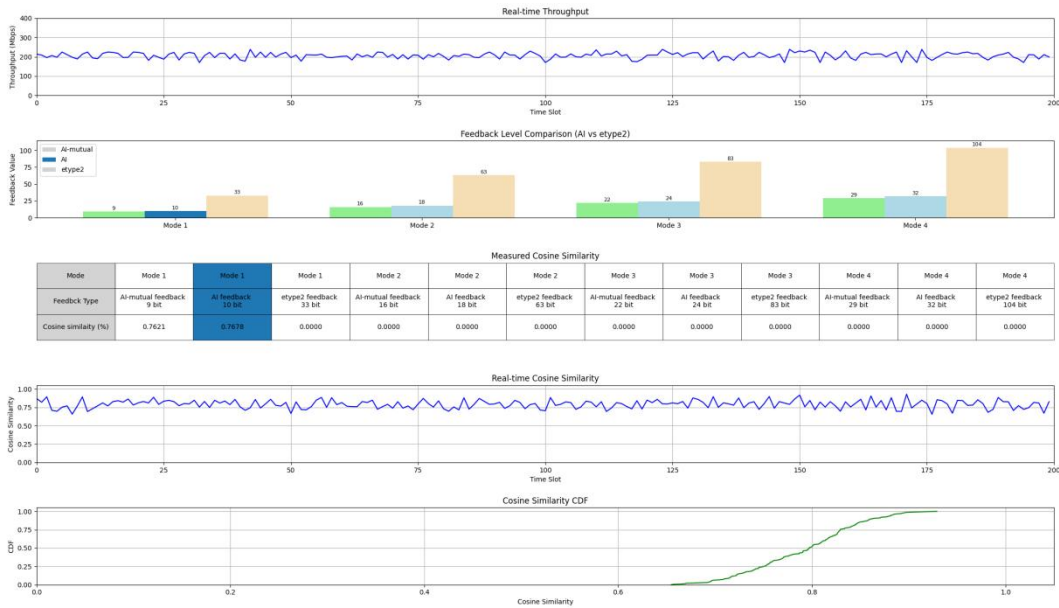


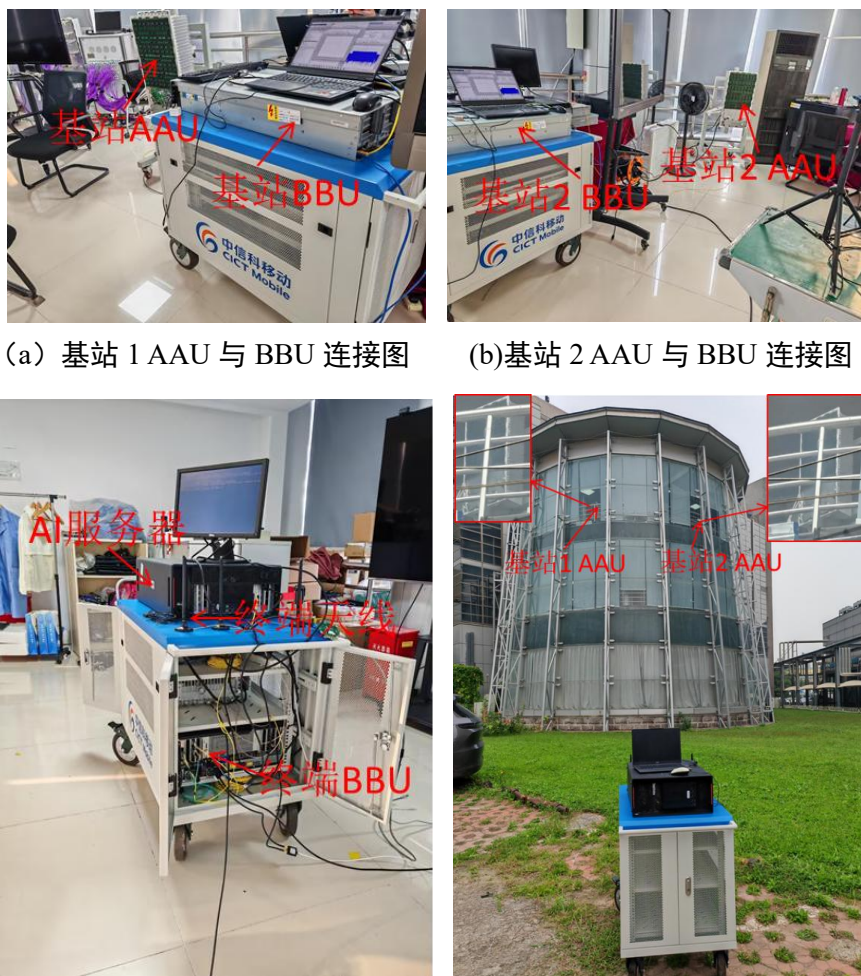
图 8-2 集中式超大规模 MIMO 样机性能结果

所提出的基于时频空信道相关性的压缩反馈方案经样机验证，在保证高吞吐量和高精度余弦相似度的前提下，降低 Etype-II 所需反馈量开销的 70%；所提出的基于上下行信道互易性的压缩反馈方案，经样机验证能够在上述基于时频空相关性方法的基础上，进一步降低所需反馈量开销的 10%。

8.2. 分布式大规模 MIMO 样机

在现有分布式 MIMO 系统中，终端通常对不同基站的 CSI 进行独立压缩与分别反馈，导致反馈开销较大。此外，各协作基站仅能获取自身与被服务终端间的 CSI，无法获知其他基站与终端之间的信道状态。白皮书提出利用分布式 MIMO 中协作基站间 CSI 的共同稀疏特性，引入基于 AI 的联合压缩与反馈机制，使基站能够同时获取其他协作基站与终端之间的 CSI 信息。

下图 8-3 展示了分布式 MIMO 下 CSI 联合压缩系统的实际部署与连接示意图，包括基站 1、基站 2、终端及 AI 服务器的物理连接方式，以及外场测试场景。测试环境中，基站部署于高楼窗侧，天线阵列朝向窗外；终端位于地面，天线指向基站方向。测试结果表明，相较于基于 3GPP Rel-16 eType II 码本的独立反馈方案，所提出的基于 AI 的联合压缩方法平均反馈精度提升 18%。



(a) 基站 1 AAU 与 BBU 连接图 (b) 基站 2 AAU 与 BBU 连接图

(c) 终端 BBU、天线、AI 服务器连接图 (d) 外场测试图

图 8-3 分布式大规模 MIMO 样机演示示意图

8.3. 超高阶调制样机

针对超高阶调制所需要面对的非线性干扰均衡问题，经过仿真验证以及原型机实测，该方案可有效补偿非线性失真，提升高阶调制传输的可靠性。在 1.3% 以上 EVM 情况下，将 4096QAM 传输的 BLER 性能降至 10% 以下，满足高阶调制传输的可靠性要求，对于 256QAM 和 1024QAM 增益同样显著。综上，该方案可在合理的 EVM 区间改善高阶调制的解调性能，跟当前业界传统方案相比，能够显著提升高阶调制的覆盖范围，提高系统谱效。

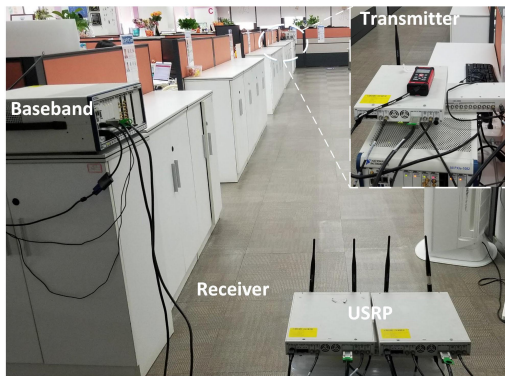


图 8-4 原型机测试环境

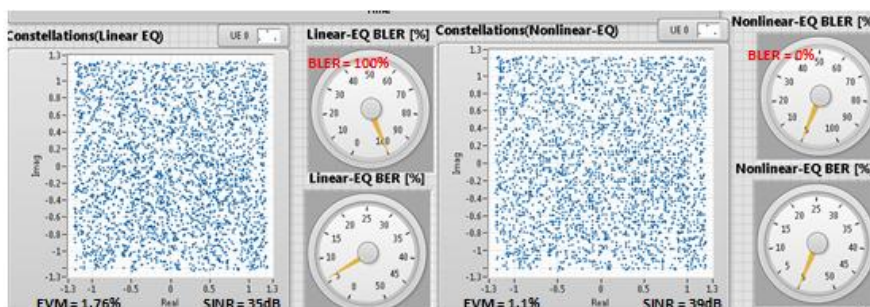


图 8-5 4096QAM 非线性均衡方案实测结果

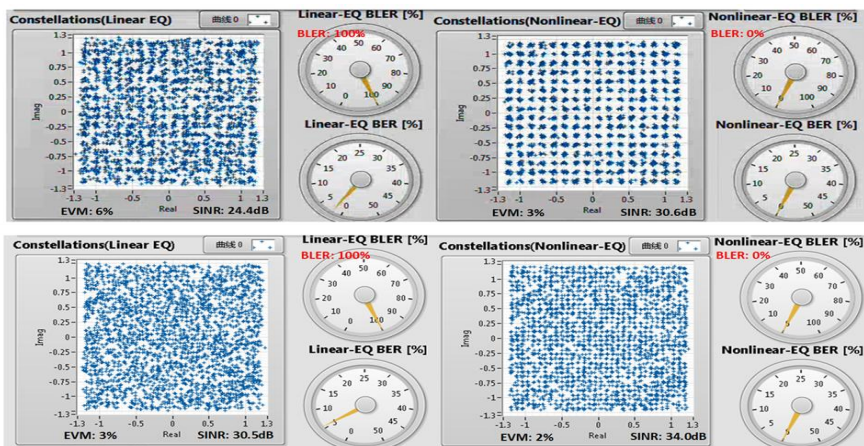


图 8-6 256QAM/1024QAM 非线性均衡方案实测结果

9. 总结与展望

将 AI 技术引入 6G 无线空口有望解决传统空口面临的谱效提升困难、基带算法复杂、大规模用户接入干扰严重等痛点问题。本白皮书立足于解决将 AI 技术引入 6G 带来的痛点问题，为业界在空口 AI 技术研发、标准化和商用落地提

供了参考。具体阐述了 6G 智能空口技术的设计思路和方法，围绕评估准则、数据集构建、智能空口传输模型设计、系统方案集成与样机等多方面提出技术创新，为 6G 无线空口迈向智能化演进提供实践路径。

展望未来，中国移动将聚焦智能空口“能力-质量-代价均衡优化”的理念，分阶段、有重点地推进智能空口技术创新、原型验证及国际标准化落地。持续攻关智能空口相关的数据集构建、模型设计等关键技术的研发与可行性验证。协同产业伙伴共建开放、协同、创新的 6G 生态，为万物智联时代奠定基石。

10. 白皮书贡献单位及人员

白皮书贡献单位包括中国移动通信有限公司研究院、北京航空航天大学、北京理工大学、清华大学、中国科学技术大学、北京三星通信技术研究有限公司、大唐移动通信设备有限公司、中关村泛联移动通信技术创新应用研究院、中山大学、上海大学。

主要贡献人员如下：

中国移动通信有限公司研究院：韩双锋、李刚、刘志明、石子焯、温子睿、李昊臻、沈嵩辉、刘继秋、张雨童、曹昱华、王菡凝、金婧、邵春菊、刘光毅

北京航空航天大学：杨晨阳、刘婷婷

北京理工大学：郑重、费泽松、张益

清华大学：孙智、王浩宇、高希钢

中国科学技术大学：赵明、周武旻、朱近康、黄杰、常江

北京三星通信技术研究有限公司：孙程君、喻斌、孙霏菲、钱辰、肖宇、白艳娜

大唐移动通信设备有限公司：索士强、高秋彬、杨现俊、黄远芳

中关村泛联移动通信技术创新应用研究院：徐飞、史文格、张淼

中山大学：王玺钧、陈翔

上海大学：孙彦赞、张舜卿、白天、朱澄煜

缩略语列表

缩略语	英文全称	中文解释
AI	Artificial Intelligence	人工智能
MIMO	Multiple Input Multiple Output	多输入多输出
3GPP	3rd Generation Partnership Project	第三代合作伙伴计划
TDD	Time Division Duplex	时分双工
FDD	Frequency Division Duplex	频分双工
CP	Cyclic Prefix	循环前缀
QoS	Quality of Service	服务质量
SVD	Singular Value Decomposition	奇异值分解
GNN	Graph Neural Network	图神经网络
ITU	International Telecommunication Union	国际电信联盟
DNN	Deep Neural Network	深度神经网络
QPSK	Quadrature Phase Shift Keying	正交相移键控
RB	Resource Block	资源块
ResNet	Residual Network	残差网络
FIFO buffer	First In, First Out buffer	先入先出存储器
CSI	Channel State Information	信道状态信息
NLOS	Non-Line of Sight	非视距
DMRS	Demodulation Reference Signal	解调参考信号
CSI-RS	Channel State Information Reference Signal	信道状态信息参考信号
PTRS	Phase Tracking Reference Signal	相位跟踪参考信号
SRS	Sounding Reference Signal	探测参考信号
OFDM	Orthogonal Frequency Division Multiplexing	正交频分复用
QAM	Quadrature Amplitude Modulation	正交振幅调制
SGCS	Squared Generalized Cosine Similarity	平方广义余弦相似度
NMSE	Normalized Mean Squared Error	归一化均方误差
AI-NC	AI-enabled nonlinearity compensation	智能非线性补偿
CNN	Convolutional Neural Network	卷积神经网络
RNN	Recurrent Neural Network	循环神经网络
ESN	Echo State Network	回声状态网络
TDL	Tapped-Delay Line	抽头延迟线
UMi	Urban Microcell	城市微蜂窝

UMa	Urban Macrocell	城市宏蜂窝
FLOPs	Floating Point Operations	浮点运算次数
GAT	Graph Attention Network	图注意力网络
FNN	Feedforward Neural Network	前馈神经网络
MAPPO	Multi-Agent Proximal Policy Optimization	多智能体近端策略优化
CDF	Cumulative Distribution Function	累积分布函数
LSTM	Long Short-Term Memory	长短期记忆网络
JSCC	Joint Source-Channel Coding	联合信源信道编码
VAE	Variational Autoencoder	变分自编码器
MG-VAE	Multinomial Gaussian-Variational Autoencoder	多项式-高斯变分自编码器
PSNR	Peak Signal-to-Noise Ratio	峰值信噪比
GFDM	Generalized Frequency Division Multiplexing	广义频分复用
ISI	Inter-Symbol Interference	符号间干扰
ICI	Inter-Carrier Interference	载波间干扰
EVM	Error Vector Magnitude	误差矢量幅度
BLER	Block Error Rate	误块率
RRC	Radio Resource Control	无线资源控制
CU	Centralized Unit	集中式单元
DU	Distributed Unit	分布式单元
CIS-RAN	Cooperative Intelligent Service-based RAN	智能协作无线接入网架构
SSCM	Scenario Specific Channel model	场景适配信道模型
CIR	Channel Impulse Response	信道冲激响应
SOTA	State-of-the-Art	前沿技术
SNR	Signal-to-Noise Ratio	信号噪声比
LS	Least Squares	最小二乘

参考文献

- [1]. 王晓云,韩双锋,刘志明,等.AI驱动的6G空口:技术应用场景与均衡设计方法[J].中国科学:信息科学,2025,55(06):1522-1533.
- [2]. Li Y, et al. Channel modeling aided dataset generation for AI-Enabled CSI Feedback: advances, challenges, and solution. *IEEE Commun. Stand. Mag.*, 2024,8(4):72-78.
- [3]. Yu B, et al. Light-weight AI enabled non-linearity compensation leveraging high order modulations. *IEEE Trans Wireless Commun*, 2024, 72(1):539-552.
- [4]. Li X, et al. Multi-task learning based CSI feedback design in multiple scenarios. *IEEE Trans Commun*, 2023, 71(12): 7039-7055.
- [5]. Wang R, et al. Proactive Hybrid Precoding for Time-varying mmWave Channel with Deep Learning. *IEEE VTC Spring*, 2023.
- [6]. Wang R, et al. Learning End-to-End Hybrid Precoding for Multi-User mmWave Mobile System With GNNs. *IEEE Trans Mach Learn Commun and Netw*, 2024, 2:978-993.
- [7]. Zhu J, et al. QAMA: Hierarchical QAM based Downlink Multiple Access with a Simple Receiver. *IEEE Transactions on Wireless Communications*, 2024, 23(10): 12565-12577.
- [8]. Chang J, et al. Distributed Joint Design of Fairness Scheduling and Beamforming in User-Dense Cell-Free Networks. *2025 IEEE VTC-Spring*, 2025.
- [9]. Sun Y, et al. Intelligent Interference Coordination Modes Selection for Clustered Cell-Free Networks. *IEEE Wireless Communications Letters*, 2025.
- [10]. Chen X, et al. Toward 6G Native-AI Network: Foundation Model-Based Cloud-Edge-End Collaboration Framework. *IEEE Commun. Mag.*, 2025, 63(8): 23-30.
- [11]. Guo Z, et al. Unsupervised AoA Estimation Based on Dual-Path Knowledge-Aware Auto-Encoders. *IEEE Trans. Wireless Commun.*, 2025, 24(5): 4425-4439.
- [12]. Bourtsoulatz E, et al. Deep joint source-channel coding for wireless image transmission. *IEEE Transactions on Cognitive Communications and Networking*, 2019, 5(3): 567-579.
- [13]. Kurka D, et al. DeepJSCC-f: Deep joint source-channel coding of images with feedback. *IEEE journal on selected areas in information theory*, 2020, 1(1): 178-193.
- [14]. Xu J, et al. Wireless image transmission using deep source channel coding with attention modules. *IEEE Transactions on Circuits and Systems for Video Technology*, 2021, 32(4): 2315-2328.
- [15]. Liu S, et al. Multidimensional Graph Neural Networks for Wireless Communications. *IEEE Trans. on Wireless Communications*. 2024, 23(4): 3057-3073.
- [16]. Zhao J, et al. Designing Heterogeneous GNNs with Desired Permutation Properties for Wireless Resource Allocation. *IEEE Open Journal of the Communications Society*, 2025, 6: 8049-8077.
- [17]. Zhao B, et al. Understanding the Performance of Learning Precoding Policies with Graph and Convolutional Neural Networks. *IEEE Trans. on Communications*, 2024, 72(9): 5657-5673.
- [18]. Liu S, et al. Learning Wideband User Scheduling and Hybrid Precoding with Graph Neural Networks. *IEEE Trans. on Wireless Communications*, Early Access,2025.
- [19]. Wang H, et al. Path Evolution Model for Endogenous Channel Digital Twin Toward 6G Wireless Networks. *IEEE Communications Magazine*,2025,63(6):34-40.
- [20]. Wang H, et al. Enhancing environment generalizability for deep learning-based CSI feedback. *arXiv preprint, arXiv:2507.06833*, 2025.

- [21]. Wang H, et al. Generalizable learning for frequency-domain channel extrapolation under distribution shift. arXiv preprint arXiv:2505.13867, 2025.
- [22]. 3GPP. TR 38.803. Study on new radio access technology: radio frequency (RF) and co-existence aspects (Release 14). 2017.
- [23]. Ayala-Romero J, et al. VrAIIn: A deep learning approach tailoring computing and radio resources in virtualized RANs. MobiCom, 2019.
- [24]. Chang H, et al. Deep echo state Q-network (DEQN) and its application in dynamic spectrum sharing for 5G and beyond. IEEE Transactions on Neural Networks and Learning Systems, 2022, 33(3): 929-939.
- [25]. Soltani M, et al. Deep learning-based channel estimation. IEEE Communications Letters, 2019, 23(4): 652-655.
- [26]. Huang T, et al. Neighbor2neighbor: Self-supervised denoising from single noisy images. IEEE/CVF conference on computer vision and pattern recognition, 2021.
- [27]. Xue Y, et al. Deep Learning-Based Joint Prediction and Compression with Optimal Prediction Interval for High-speed CSI Feedback. IEEE VTC Fall Workshop, 2025.
- [28]. Wang K, et al. AI-Enhanced CSI Feedback via Exploiting Multi-user Shared Information in mMIMO Systems. IEEE VTC Fall Workshop, 2025.
- [29]. Jia M, et al. Deep learning based unified CSI feedback for TDD and FDD massive MIMO systems. IEEE VTC Fall Workshop, 2025.11.14.
- [30]. Wang X, et al. C-ran advanced: from a network cooperation perspective. arXiv preprint arXiv: 2511.03302, 2025.