arXiv:2603.02745v1 [cs.IT] 3 Mar 2026

采用深度强化学习方法提升多面板毫米波无线接入网中基于波束的多用户多输入多输出用户吞吐量

Ramin Hashemi†、Vismika Ranasinghe†、Teemu Veijalainen†、Petteri Kela† 和 Risto Wichman‡

†诺基亚，芬兰埃斯波，电子邮箱：{ramin.hashemi, vismika.ranasinghe, teemu.veijalainen, petteri.kela}@nokia.com

‡阿尔托大学信息与通信工程系，芬兰埃斯波，电子邮箱：risto.wichman@aalto.fi

摘要——**毫米波（mmWave）通信系统，尤其是采用混合波束成形的多用户多输入多输出（多用户多输入多输出）系统，由于动态波束选择和管理的复杂度很高，在优化用户吞吐量和降低时延方面面临挑战。本文提出一种深度强化学习方法，用于在实际网络设置中提升多面板毫米波无线接入网的用户吞吐量。我们基于深度强化学习构建的方案采用自适应波束管理策略，将通信智能体与其环境之间的交互建模为马尔可夫决策过程（MDP），并根据实时观测优化波束选择。所提出的框架利用空间域（SD）特性，纳入不同天线面板中波束之间的互相关、测得的参考信号接收功率（RSRP）以及波束使用统计，以动态调整波束成形决策。因此，该方法提升了频谱效率并降低了端到端时延。数值结果表明，与基线（传统波束管理）相比，吞吐量最高提升 16%，时延降低至原来的 1/3–1/7。**

索引词——**基于波束的多用户多输入多输出、深度强化学习（深度强化学习）、混合波束成形（HBF）、毫米波通信。**

I. 引言

混合波束成形是一种用于无线通信系统的技术，尤其适用于大规模多输入多输出（MIMO）系统，可在射频链数量有限的情况下，高效权衡性能与硬件复杂度[1]，[2]。它结合了数字波束成形和模拟波束成形。在混合波束成形中，使用数量较少的射频链，从而能够先进行多流数字基带处理，再进行模拟处理，以实现天线波束成形增益。这种方法尤其适用于毫米波通信，因为全数字波束成形的成本和功耗高得难以承受 [3]。此外，它还能通过降低波束成形所需的功耗实现节能运行，使系统能够利用不同波束同时为多个用户提供服务。

基于波束的多用户多输入多输出（MU-多输入多输出）[1] 的核心原理是利用无线信道的空间维度（SD）来提升容量。通过采用先进的信号处理算法，系统能够确定最优波束成形向量，在最大化目标用户所在位置的信号强度的同时，尽量减少对其他用户的干扰。这需要借助信道状态信息（CSI）来实现。近期研究广泛探讨了毫米波和大规模多输入多输出系统中的混合波束成形与用户调度。在硬件约束下采用贪心用户调度等经典方法已在[4]，中得到研究，而面向宽带毫米波正交频分复用（OFDM）的混合模拟数字波束成形设计则在 [5]中提出。深度学习也日益广泛地应用于降低信道估计和混合波束成形的复杂度 [6], [7]。近期研究还探讨了不完美信道状态信息下的调度，以及 OFDM 混合波束成形中的复杂度问题[8]，，并提出了深度学习增强型混合波束成形策略 [9]。与这些研究相比， [10]中基于学习的比例公平调度框架侧重于双时间尺度协议，并明确量化了公平性与计算复杂度之间的权衡。然而，采用监督学习技术需要收集海量数据集，这在实际网络环境中尤其难以实现，而且模型难以适应不同的无线环境。因此，在此类场景中，利用通过与无线环境交互来学习策略的强化学习（RL）显然具有优势。

在我们先前的工作 [11]，中，我们研究了一种由强化学习辅助的波束选择方法，用于毫米波无线接入网中的模拟波束成形应用场景。研究考虑了单天线面板 gNB（仅配备一条射频链），这使波束选择问题简化为两个维度，即利用波束使用统计信息和测得的参考信号接收功率值来选择波束。然而，在多面板 gNB 中必须考虑三个维度，因为多个波束会同时

[1]请注意，基于波束的 MU-MIMO 框架由混合波束成形架构支持，可通过多个天线面板同时向多个方向传输信号。

启用，并可能相互干扰。换言之，波束网格（GoB）的空间域特性与上述维度共同构成了波束选择所涉及的维度。因此，本研究的主要创新在于提出了一种波束管理框架，在采用多个天线面板的基于波束的多用户多输入多输出技术中，综合利用波束使用（或波束流行度）、测得的参考信号接收功率值以及空间域互相关。鉴于所提出的吞吐量优化问题复杂度很高，我们采用强化学习方法来解决这一问题。

I. 系统模型与问题表述

我们考虑由 M 个基站（gNB）组成的蜂窝网络的下行链路（DL），这些基站协同服务于 U 个移动终端（MT），所有 MT 的集合记为 $\mathrm{U} = \{1, 2, \cdots, U\}$。如图1所示，每个 gNB 配备 M_p 个天线面板，每个面板均采用均匀方形平面阵列（USPA）实现。为进行定向传输，每个面板基于波束网格结构支持预定义的 L_p 个模拟波束。因此，所有天线面板上的波束总数为 $B = M_p \times L_p$。为简单起见，我们将所有面板上的波束汇集为集合 $\mathrm{B} = \{1, 2, \ldots, B\}$，其中每个波束 $b \in \mathrm{B}$ 可由面板 p 和局部索引 $j \in \{1, 2, \ldots, L_p\}$ 标识，使得 $b = pL_p + j$。系统采用模拟波束成形架构，每个面板配备一条射频链。这意味着在一个传输时间间隔（TTI）内，每个面板只能激活一个波束。因此，最多可同时激活 M_p 个波束。移动终端均匀分布在服务区域内，并按照预定义的速度遵循相应的移动性配置。无线传播环境采用标准化的 3GPP 三维空间信道模型 [12] 进行建模。模型涵盖了所有必要的 5G新空口（NR）功能，包括物理层（PHY）过程、媒体访问控制（MAC）以及波束管理机制。按照 3GPP 规范，每个 gNB 在时长为 T_{bs}（例如，40 ms）的波束切换周期内，对每个面板的所有候选波束执行波束扫描。移动终端测量各波束的参考信号接收功率（RSRP），并将部分测量结果反馈给服务 gNB，由其确定激活的波束 [13]。

gNB处的多面板配置增强了空间覆盖和分集，但也带来了额外的复杂性进行波束选择。由于每个面板仅配备一条射频链，因此每个传输时间间隔内每个面板只能调度一个波束。因此，如果所选波束与移动终端业务需求不匹配，资源块就可能未得到充分利用。如图1所示，仅依据最强参考信号接收功率选择波束，并不一定能最大化频谱效率，因为图2展示了三个维度。在某些情况下，参考信号接收功率稍弱但调度频率更高的波束，能够更有效地利用可用带宽。此外，若不了解互相关值而选择波束，也无法实现吞吐量优化。这一现象揭示了

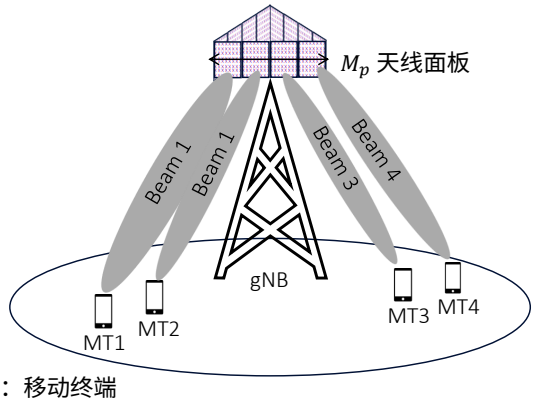

图1：采用 $M_p = 4$ 天线面板的基于波束的多用户多输入多输出中的波束选择方法。每个面板中激活的波束与其他面板中的波束在空间上彼此分离；然而，空间距离较近的波束之间可能存在非零交叉相关性。

一个关键权衡：选择参考信号接收功率更强的波束以最大化瞬时链路质量，与利用激活概率更高或互相关值更低的波束以提升长期效率之间的权衡。在多面板部署中，必须谨慎优化这一权衡，因为各面板间的波束选择决策相互依赖。因此，需要先进算法综合考虑信号强度、历史激活统计和面板间波束动态，以实现系统级性能提升。

如图所示。 2 在基于波束的多用户多输入多输出应用场景中，波束选择准则需要考虑若干新增维度。为了优化系统性能、提高吞吐量和链路质量（例如信号与干扰加噪声比），可以通过频分复用将多个移动终端聚类到同一波束上，从而减少移动终端之间的频域资源干扰，帮助缓解干扰。此外，使用能获得更多调度机会的热门波束，可以避免缓冲数据包出现不必要的延迟（因为激活新波束可能导致延迟）。另外，高效地分隔波束可实现空分复用，使两个移动终端能够通过不同波束进行共同调度。此方法利用空间分隔来减少交叉干扰，从而最大限度地提高吞吐量/信号与干扰加噪声比。

我们用 $\rho_{b,j}, b \neq j$ 表示不同面板内各波束之间的互相关值，分别对应波束 b和 j。已知 $\rho_{b,j}$ 会显著影响传输质量，并最终显著影响所实现的吞吐量。为从统计上计算该值，首先确定在过去的调度间隔（即过去的传输时间间隔）中先前分配给移动终端的波束集合。对于每个波束，按每个移动终端跟踪其历史激活次数。任何至少被选择过一次的波束都被归类为潜在干扰波束，并纳入干扰候选集。因此，源自不同天线面板的各波束之间的空间互相关通过下述过程计算：

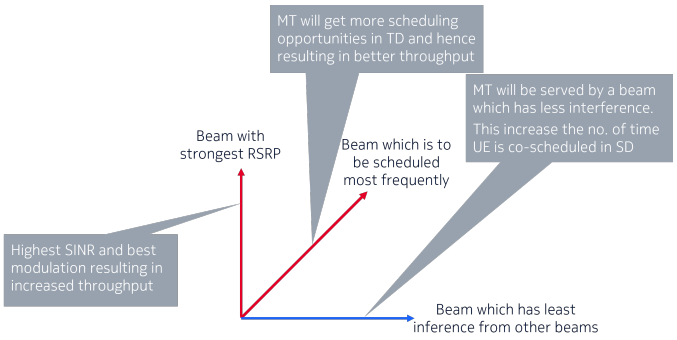

图2：基于波束的多用户多输入多输出波束管理流程中，除传统的基于参考信号接收功率的经典方法外还需考虑的可能维度。

- 对于第一个移动终端（具有最高调度优先级），其波束被预测时，向量的互相关部分全部为零。
- 对于首个移动终端之后的每个移动终端：– 找出已分配给当前正在进行波束选择的移动终端的波束。– 对于当前移动终端考虑的所有候选波束，计算其与已分配给其他移动终端的波束之间的互相关之和。

在增强型移动宽带（eMBB）场景中，一个重要的设计目标是最大化成功传输至移动终端的数据量。我们将其表述为单用户最大化问题，并假设调度器负责管理用户正交性。该要求可表示为以下优化问题：

$$\mathbf{P1:}\ \max_{b}\ \Gamma_u^{\mathrm{tot}} = \sum_{t=1}^{T} \gamma_t^u(h_b, z_b', \rho_{b,j}) \tag{1}$$

$$\text{约束条件 } \mathrm{C}_1:\ -140 \le z_b' \le -44, \quad \forall b \in \mathrm{B},$$

$$\mathrm{C}_2:\ 0 \le h_b \le \frac{T_{\mathrm{bs}}}{T_{\mathrm{TTI}}}, \quad \forall b \in \mathrm{B},$$

$$\mathrm{C}_3:\ \rho_{b,\,j} \in [0,1], \quad \forall b,\ \ j \in \mathcal{B}\ \ b \neq j,$$

其中 T 是一个波束切换间隔内传输时间间隔的总数 T_{bs}，Γ_u^{tot}表示用户 u在一个波束切换间隔内的总吞吐量，而 $\gamma_t^u(h_b,\ z_b',\ \rho_{b,\ j})$ 是在第 t 个传输时间间隔内实现的吞吐量，记为 T_{TTI} 其受许多因素影响，例如无线信道增益、干扰功率以及 gNB 的调度器算法 [14]。约束 C_1 将原始参考信号接收功率测量值 z_b' 限制在 3GPP 规定的报告范围内。约束 C_2 限制激活计数器 h_b，该计数器记录在该间隔内波束 b被选中的次数，其上限为传输时间间隔数，即波束切换间隔内的 $T_{\mathrm{bs}}/T_{\mathrm{TTI}}$ T_{bs}。约束 C_3 是互相关系数的上限$\rho_{b,\ j}$用于表示波束 b和 j共同调度的概率。

问题 **P1** 表明，有效的波束选择需要权衡三个因素：较强的信号强度（z_b）、过去的波束使用情况（h_b）以及波束间互相关（$\rho_{b,\ j}$）。随着波束数量 B 的增加，直接优化的计算复杂度极高，在实践中不可行。强化学习提供了一种可扩展的替代方案：利用状态特征 $\{h_b,\ z_b,\ \rho_{b,\ j}\}$ 训练智能体，使其学习选择策略。训练完成后，通过策略推理即可获得决策，无需进行复杂的迭代优化。

III. 基于深度强化学习的建模

在强化学习辅助的波束管理中，智能体旨在确定一种最优策略 π^* ：S → A，通过将观测到的信道状态映射为波束成形动作，最大化预期通信性能。智能体与无线环境之间的交互被建模为马尔可夫决策过程（MDP），定义为 (S，A，R，P)，其中：

- S 表示状态空间，可能包括信道状态信息系数、参考信号接收功率值或之前的波束索引，
- A 表示离散动作空间，对应于可用波束索引的集合，
- R ：S ×A → ℝ 定义奖励函数，该函数可以设计为瞬时频谱效率、信号与干扰加噪声比（SINR）或可实现吞吐量，
- P($s\prime|s,\ a$) 指定在执行动作 a ∈A 时，从状态 s ∈ S 到达状态 $s\prime$ ∈ S 的状态转移概率。

智能体的目标是最大化期望折扣累积奖励 $J(\pi) = \mathbb{E}_\pi\left[\sum_{t=0}^{\infty}\gamma^t r_t\right]$，其中 $\gamma \in [0,1)$ 为折扣因子，r_t 表示时间步 t 的即时通信奖励（例如频谱效率）。策略 π 下的动作价值函数为

$$Q^\pi(s,a) = \mathbb{E}_\pi\left[\sum_{t=0}^{\infty}\gamma^t r_t \,\middle|\, s_0 = s, a_0 = a\right], \tag{2}$$

其满足贝尔曼期望方程：

$$Q^\pi(s,a) = \mathcal{R}(s,a) + \gamma \sum_{s'\in\mathcal{S}} \mathcal{P}(s'|s,a)\, \mathbb{E}_{a'\sim\pi(\cdot|s')}[Q^\pi(s',a')]. \tag{3}$$

最优波束选择策略由最优动作价值函数 $Q^*(s,a)$ 导出，该函数满足贝尔曼最优方程：

$$Q^*(s,a) = \mathcal{R}(s,a) + \gamma \sum_{s'\in\mathcal{S}} \mathcal{P}(s'|s,a) \max_{a'\in\mathcal{A}} Q^*(s',a'). \tag{4}$$

在实际的毫米波和太赫兹通信场景中，由于信道稀疏性、用户移动性和硬件损伤，精确建模状态转移动态 P($s'|s,a$) 是不可行的。为克服这一限制，采用无模型强化学习算法，直接近似 $Q^*(s,a)$ 或其对应策略，而无需显式了解 P。该表述使自适应波束选择成为可能，智能体根据观测到的 RSRP测量值以及与特定于 gNB 的信息相对应的性能反馈，逐步优化其波束成形决策。

假设强化学习阶段考虑了 K 个候选波束，原始 RSRP测量值为 $\mathbf{z}' = [z_1', z_2', \ldots, z_B']$ 其中

z'_b 对应于第 k 个波束测得的 RSRP。根据 3GPP 规范，这些测量值会被截断至 $[-140, -44]$ dBm 的范围内，然后归一化为

$$\mathbf{z} = \frac{\mathbf{z}' + 140}{96}, \quad \mathbf{z} = [z_1, z_2, \ldots, z_B]. \tag{5}$$

此外，波束激活历史表示为

$$\mathbf{h} = [h_1, h_2, \ldots, h_B], \tag{6}$$

其中，h_b 表示最近一个波束切换间隔内第 k个波束的归一化使用计数器。每个间隔约为 40 ms，其中包含多个传输时间间隔。在每个传输时间间隔内，调度器会选择一个活动波束，并将各传输时间间隔内的累计使用统计信息存储在$\mathbf{h}$中。

波束选择任务通过在强化学习框架中定义状态、动作和奖励函数来处理。我们将对其进行详细探讨：

1）状态：强化学习智能体通过构建融合多个维度信息的状态表示，形成决策所需的上下文。与我们之前的工作类似[11], , 状态向量包含：(i) 候选波束的归一化参考信号接收功率值，(ii) 前一个切换间隔内的波束激活频率，以及 (iii) 活动移动终端的服务波束之间的成对互相关指标，这些指标反映了它们被联合调度的可能性。

因此，时间步 t 的状态为

$$\mathbf{s}_t = \left\{ h_b,\ z_b,\ \rho_{b,j} \,\middle|\, \forall b, j \in \{1, 2, \ldots, B\},\ b \neq j \right\}, \tag{7}$$

其中，z_b 表示波束 b 的归一化参考信号接收功率，h_b 表示波束 b 的归一化激活频率，而 $\rho_{b,\ j}$则反映不同天线面板内波束 b 和 j 之间的互相关。之所以纳入激活统计数据而非缓冲区相关流量描述符，是因为过去的波束使用模式能够更可靠地指示未来的调度倾向。状态空间的总体维度随波束数量 B 及互相关项的数量而增大。

2）动作：在每个决策时刻，智能体从可用候选项集合中选择一个波束。策略网络针对 B 个波束索引输出动作价值，并通过选取输出中的最大值索引来确定动作。时间步 t的动作a_t定义为 $a_t \in \{1, 2, \ldots\ldots, B\}$，其中 $a_t = b$ 表示在时隙 t内激活第 b个波束。

3）奖励：设 B_k^t 表示在时间步 t成功传送给用户 k 的数据字节数。奖励函数基于移动终端的观测吞吐量。此外，为了稳定训练过程，奖励按照所有用户中传送的数据字节数最大值进行归一化。分配给智能体的奖励为

$$r_t = \frac{B_k^t}{\max_j(B_j^t)}, \tag{8}$$

其中，$\max_j(B_j^t)$ 表示时间 t 所有活跃用户中的峰值传输结果。这种归一化可防止奖励幅度变得无界，并加快学习算法 [15] 的收敛。

表I：仿真配置。

参数	值
部署场景	3GPP 密集城区宏蜂窝 [17], [12] 中的三维信道模型
gNB 发射功率	40 dBm
移动终端数量 (U)	210
gNB 数量 (M)	21（7 个站点，每个站点 3 个扇区）
每个 gNB 的天线面板数 (M_p)	每个扇区 3 个面板
每个面板的波束数（L_p）	16 个波束
波束总数（B）	$B = L_p \times M_p = 48$
波束切换间隔（T_{int}）	40 毫秒
混合波束成形算法	相关性阈值
MT 配对相关性阈值	0.4
MT 速度	3 km/h
载波频率	FR2，30 GHz
系统带宽（W）	200 MHz
子载波间隔	60 KHz
接收机噪声系数	3 dB
热噪声功率谱密度（N_0）	−174 dBm/Hz
业务模型	FTP3，21 兆比特每秒目标速率
分组大小	600 字节
训练批次大小	32 个样本
经验回放大小	5000
学习率（Adam优化器）	10^{-4}
激活函数	$ReLU(\cdot)$
神经网络架构	（输入，128，256，输出）
训练轮次	4

在基于波束的多用户多输入多输出中，波束管理框架内的最优波束选择策略通过双重深度 Q 网络（双重深度 Q 网络）方法获得，该方法用于近似动作值函数并确定使预期Q值最大的波束 [15]。智能体使用的状态表示由归一化的RSRP测量值、波束激活历史和互相关值组成，如上所述。

IV. 数值结果

在本节中，我们展示系统级仿真结果。我们使用真实的系统仿真平台来测试所提出的波束选择方法[2]。智能体的目标是在 FTP3 业务模型下最大化下行链路总吞吐量，其中每个移动终端的业务负载为 21 兆比特每秒。除非另有说明，我们假设该模型位于 gNB 侧。我们采用宏蜂窝场景，其中由标准六边形网格排列的三扇区 5G NR gNB 组成，站间距为 200 米；载波频率为 30 GHz，符合 3GPP 的 FR2 指南。表 I 总结了仿真所采用的重要参数。为在多个天线面板上实现基于波束的多用户多输入多输出，我们采用混合波束成形框架，根据互相关阈值对移动终端进行配对，以最大化时频域中的比例公平性指标。在此方法中，若移动终端满足相关性阈值条件，就会被选中与第一个移动终端配对。基于强化学习的波束选择的作用是在执行混合波束成形之前，调整与各移动终端相关联的波束。

训练期间的探索性波束选择会影响性能，可能导致性能暂时低于基线水平，如图所示。 3。由于强化学习训练具有闭环特性，因此需要进行探索，以学习

[2]有关所用仿真引擎的详细信息，请参阅 [13], [16]。

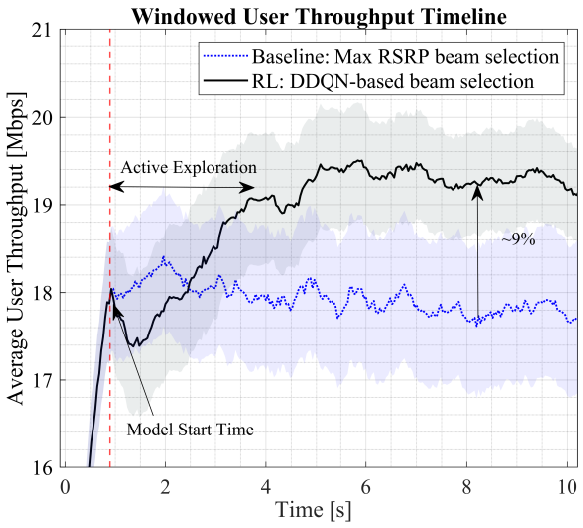

图 3：训练期间探索对移动终端平均吞吐量的影响（网络中的移动终端总数为 $U = 210$）。

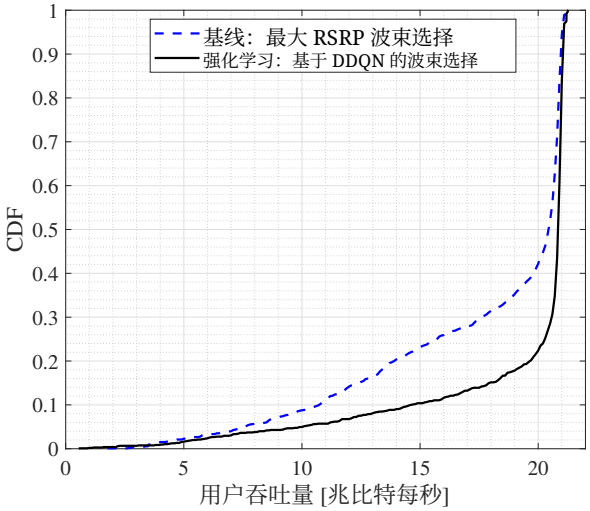

图 4：用户吞吐量的累积分布函数（$U = 210$）。

所选波束与观测奖励（即吞吐量）之间的关系。图 3 展示了探索各种波束对移动终端平均吞吐量随时间变化的影响（网络中有 210 个移动终端）。在主动探索期间，平均吞吐量比模型收敛后的平均吞吐量低约 10%，并低于基线方法。

图 4 展示了由 210 个移动终端组成的网络中用户吞吐量的累积分布函数（CDF）。该图比较了两种波束选择策略：采用最大 RSRP 进行波束选择的基线方法，以及采用 DDQN 进行波束选择的强化学习方法。CDF 曲线表明，基于 DDQN 的强化学习波束选择始终优于基线方法，使整个分布中的用户都能获得更高的吞吐量。具体而言，强化学习方法的 CDF 曲线更陡，说明与基线方法相比，更大比例的用户能够获得较高吞吐量。这表明基于强化学习的方法能够有效优化网络性能并提升用户体验。

图 5 比较了基于强化学习的波束选择与最强参考信号接收功率波束选择方法（即基线，也就是传统方法）。如图所示，基于强化学习的框架能够带来更高的总体／有效用户吞吐量几何平均值。有效吞吐量的计算方法是：移动终端接收的数据总量除以其被调度接收的总时间。此外，总体吞吐量的计算方法是：移动终端接收的数据总量除以其处于活动状态的总时间。为了证明所有移动终端都实现了吞吐量优化，我们将用户吞吐量几何平均值与基线波束选择进行比较。几何平均值（GM）定义为

$$\mathrm{GM} = \sqrt[U]{\Gamma_1^{\mathrm{tot}}\Gamma_2^{\mathrm{tot}}\cdots\Gamma_U^{\mathrm{tot}}}, \quad \Gamma_u^{\mathrm{tot}} \neq 0 \quad \forall u \in \mathcal{U} \tag{9}$$

其中，Γ_u^{tot} 是移动终端u的计算吞吐量值。此外，由于优化了波束选择并改进了通过强化学习为移动终端提供调度机会，端到端时延已大幅降低。具体而言，与基线（即最大参考信号接收功率波束选择方法）相比，基于强化学习的波束选择优化可带来高达 5–16% 的吞吐量增益，并使端到端时延改善 3–7 倍。时延改善的主要原因是强化学习智能体学习到的策略。该智能体会调度并分组使用次优波束（就参考信号接收功率而言）的移动终端，使 gNB 能够立即分配资源。因此，用户的数据包无需在等待特定波束激活时长时间缓冲。

强化学习智能体利用波束的空间、时间和接收功率特征来学习该策略。例如，如图 6 所示，强化学习智能体分组的移动终端数量多于基线，从而增加了调度机会。请注意，基于波束的多用户多输入多输出中，共同调度用户的最大数量始终受天线面板数量限制，即 $M_p = 3$。

五、结论

本研究提出了一种基于深度强化学习的多面板毫米波无线接入网波束管理框架，旨在提高用户吞吐量并降低时延。我们采用可扩展的强化学习方法，解决了如**P1**所述多用户多输入多输出系统中优化波束选择涉及的复杂问题，而传统优化方法难以胜任。与基准波束选择技术相比，我们提出的方案显著提升了吞吐量，最高增幅达 16%，并将端到端时延大幅降低至原来的 1/3–1/7。这些发现证实了强化学习在动态无线环境中的有效性，也为管理复杂网络场景、提升毫米波频段的用户体验提供了有益见解。未来研究可进一步探索优化方法，并纳入更多维度的信道状态信息，或采用数字波束成形技术。

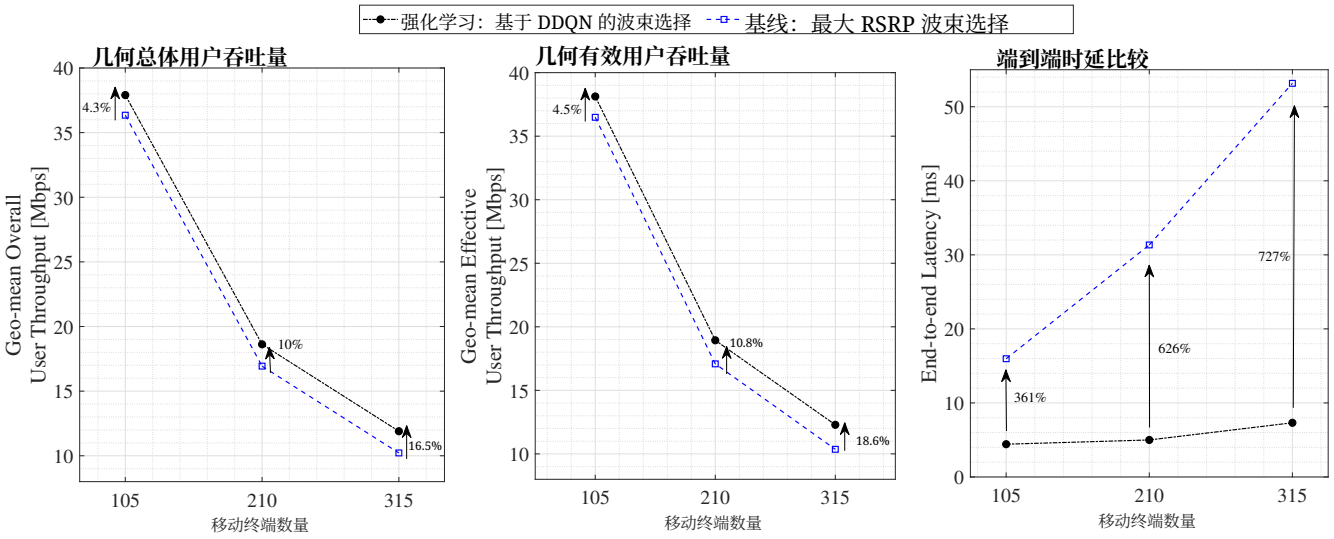

图 5：比较 RL-DDQN 与基线的用户吞吐量几何平均值和端到端时延。

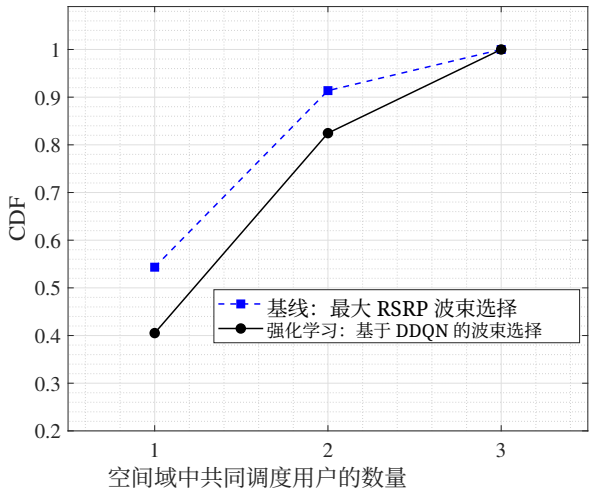

图 6：空间域中共同调度用户的累积分布函数。

参考文献

[1] I. Ahmed、H. Khammari、A. Shahid 等，“5G 混合波束成形技术综述：架构与系统模型视角”，*IEEE Communications Surveys & Tutorials*，第 20 卷，第 4 期，第 3060–3097 页，2018 年。[2] J. Kim 和 M. Andrews，“毫米波混合波束成形系统中基于学习的自适应用户选择”，载于*ICC 2023 - IEEE* 国际通信会议，意大利罗马，2023 年，第 5645–5650 页。[3] S. Sun、T. S. Rappaport、R. W. Heath 等，“面向毫米波无线通信的多输入多输出：波束成形、空间复用，还是两者兼有？”，*IEEE Communications Magazine*，第 52 卷，第 12 期，第 110–121 页，2014 年。[4] J. Choi、G. Lee 和 B. L. Evans，“采用低分辨率模数转换器的毫米波混合波束成形系统中的用户调度”，*IEEE Transactions on Wireless Communications*，2018 年。[5] F. Sohrabi 和 W. Yu，“面向毫米波正交频分复用大规模天线阵列的混合模拟与数字波束成形”，*IEEE Journal of Selected Topics in Signal Processing*，2017 年。

[6] A. M. Elbir 等，“用于多载波毫米波大规模多输入多输出系统信道估计和混合波束成形的一系列深度学习架构”，IEEE Transactions on Signal Processing，2019。[7] Z. Gao 等，“基于数据驱动深度学习的混合波束成形，用于具有隐式信道状态信息的空中大规模 MIMO-OFDM 系统”，IEEE Transactions on Communications，2022。[8] J. Liu 和 B. Wang，“非理想信道状态信息下毫米波混合波束成形 OFDM 系统的低复杂度用户调度框架”，Recent Advances in Computer Science and Communications，第 18 卷，第 2 期，2025。[9] M. A. A. Abir、M. Foysal、A. Hossan 等，“在多输入多输出系统中使用接收信号强度指示信号实现增强型混合波束成形的深度学习方法”，e-Prime - Advances in Electrical Engineering, Electronics and Energy，第 9 卷，第 100738 页，2024。[10] P. Agheli、T. Kobal、F. Durand 等，“采用混合波束成形的 MIMO-OFDM 系统中基于学习的多用户调度”，载于 2025 年欧洲网络与通信联合会议暨 6G 峰会（EuCNC/6G 峰会），波兹南，波兰，2025 年，第 1–6 页。[11] R. Hashemi、V. Ranasinghe、T. Veijalainen 等，“通过深度强化学习优化毫米波无线接入网中的波束切换用户吞吐量”，载于 2024 年第 35 届 IEEE 个人、室内和移动无线电通信国际研讨会（PIMRC），瓦伦西亚，西班牙，2024 年，第 1–6 页。[12] 3GPP，“0.5 至 100 GHz 频段的信道模型研究”，3GPP TR 38.901。 [在线]。可访问：https://www.3gpp.org/ dynareport/38901.htm [13] F. Abinader、C. Rom、K. Pedersen 等，“采用不同多面板天线设备模型的毫米波 5G 系统级分析”，载于 2021 年第 93 届 IEEE 车辆技术会议（VTC2021-Spring），赫尔辛基，芬兰，2021 年，第 1–6 页。[14] P. Kela、B. Liu 和 A. V. Rial，“从仿真到实践：面向蜂窝调度器的可泛化深度强化学习”，载于 NeurIPS 2025 研讨会：面向下一代无线通信与网络的人工智能和机器学习，2025。[15] R. S. Sutton、A. G. Barto 等，《强化学习：导论》，麻省理工学院出版社，剑桥，1998 年，第 1 卷，第 1 期。[16] K. Pedersen、R. Maldonado、G. Pocovi 等，“无线系统级仿真教程：侧重 3GPP 5G-Advanced 及后续技术”，IEEE Communications Surveys & Tutorials，第 26 卷，第 4 期，第 2290–2325 页，2024。[17] 3GPP，“下一代接入技术的场景与需求研究”，3GPP TR 38.913。 [在线]。可访问：https: // www.3gpp.org/dynareport/38913.htm