XXXX

基于Transformer-双Q网络的太赫兹NOMA通信网络即时功率分配

周音

（中国人民解放军91977部队，北京 100036）

摘　要：太赫兹非正交多址接入（non-orthogonal multiple access，NOMA）技术有望成为6G通信系统的关键突破性方案。其核心机制是通过利用超宽带资源与功率域复用，实现海量用户共享多个子频段，显著提升系统连接容量。为充分释放太赫兹NOMA系统的性能潜力，关键在于实现满足服务质量（quality of service，QoS）约束下的快速功率分配优化。提出了一种基于Transformer架构的双Q网络模型，通过Transformer学习不同用户分配策略的关联性，并采用双Q网络实现更稳定的决策优化过程。经训练后本算法可生成适应多种用户分布的即时功率分配策略。实验结果表明，训练完成的模型仅需毫秒级计算时间，即可实现接近穷举法的高吞吐量性能。该算法展现出较强的实时性与鲁棒性，具有较大工程应用潜力。

关键词：太赫兹通信；非正交多址网络；资源分配

中图分类号：TN92

文献标志码：A

doi: 10.11959/j.issn.1000−0801.

Instant power allocation for terahertz NOMA communication networks based on transformer-double deep Q network

ZHOU Yin

Unit 91977 of the PLA, Beijing, 100036, China

Abstract: Terahertz (THz) non-orthogonal multiple access (NOMA) was regarded as a candidate in 6G and beyond systems. By exploring the ultrabroad bandwidth and power domain, THz-NOMA could realize massive connectivity through assigning each sub-band to different users. To unleash the potential of the THz-NOMA system, it was significant to allocate power fast under quality of service (QoS) requirements. Focusing on the instant power allocation, a novel transformer-based double deep Q-network (DQN) solution adaptive for general user distributions was proposed in this paper. Transformer was used to learn the relationships among allocation strategies for different users, and a double DQN was adopted to achieve a more stable decision optimization process. The simulation results validated that the proposed algorithm realized the throughput close to the optimum given by exhaustive search method within millisecond level. The proposed method demonstrates high real-time performance and robustness, which suggests its high practicability.

收稿日期：XXXX−XX−XX；修回日期：XXXX−XX−XX

Key words: terahertz communication, non-orthogonal multiple access, resource allocation

0 引言

随着移动互联网和智能终端的迅猛发展，人类社会对信息传输的依赖程度不断加深，视频流媒体、虚拟现实、物联网以及自动驾驶等新兴应用持续涌现，这些应用均对无线通信系统提出了更高要求。未来无线通信不仅需要更高的数据速率和更低的时延，还需具备更高的频谱利用效率以及支持海量终端同时接入的能力[1][2]。然而，传统无线通信主要依赖的微波与毫米波频段频谱资源日趋紧张，使得传统无线接入技术在满足超高速率与超大规模连接需求方面面临瓶颈[3]。太赫兹通信作为6G潜在的关键技术方向之一，有极为丰富的可用频谱资源（0.1 – 10 THz），同时具备高带宽和高方向性的传播特性，在实现未来高速、大容量通信方面展现出巨大潜力[4][5][6]。此同时，非正交多址接入（non-orthogonal multiple access，NOMA）技术通过功率域复用方式，使多个用户能够在相同时间以及频率资源上以不同功率同时传输信号，从而突破传统正交多址接入的限制，显著提高频谱效率和系统容量，被视为支持大规模连接和高频谱效率的重要技术手段[7][8][9]。将太赫兹通信与NOMA技术相结合构建太赫兹NOMA网络，有望在高频带宽优势与高效多址接入机制之间形成协同效应，为未来无线通信系统提供更加高效的解决方案。

在太赫兹NOMA系统中，功率分配策略是影响系统性能的关键因素之一，它直接决定了不同用户之间的干扰关系、信号解码顺序以及系统整体吞吐量。合理的功率分配不仅可以提升系统容量，还能兼顾用户公平性与通信可靠性。为了充分发挥NOMA技术的优势，研究者提出了多种优化方法，例如将多输入多输出（multiple-input multiple-output，MIMO）技术与NOMA结合，通过空间复用与功率域复用的联合设计进一步提升数据传输速率与系统容量[8][10]。此外，文献[11]提出联合波束成形、功率分配以及带宽分配的综合优化方案，通过迭代算法不断逼近穷举搜索得到的最优吞吐量，从而在理论上实现更优的系统性能。这些研究为提升太赫兹NOMA系统的资源利用效率提供了重要参考。

然而，上述基于迭代优化或穷举搜索的方法通常计算复杂度较高，尤其是在用户数量增多或信道状态快速变化的情况下，算法计算耗时显著增加，难以满足实际无线通信系统对实时性和低延迟调度的要求。近年来为解决这一问题，机器学习方法被引入资源分配领域[12][13][14]。经过训练的机器学习模型能够根据当前信道状态快速生成资源分配策略，从而避免重复执行复杂优化计算。针对NOMA系统功率分配问题，文献[15]提出基于有标签深度学习的方法，通过利用最优分配结果作为训练标签来学习映射关系，但该方法依赖高质量最优标签，而这些标签往往需要复杂优化算法生成，获取成本较高。文献[16]采用Q学习方法进行长期功率分配决策，但其需要计算每个状态-动作对对应的速率，在状态空间和动作空间较大的情况下计算负担较重。文献[17]采用深度Q网络，通过引入深度神经网络逼近Q函数，无需预先获得最优标签即可学习长期最优策略，但该方法更关注长期平均性能，难以确保每个时刻的瞬时性能最大化，而且模型往往针对训练环境进行优化，泛化能力有限，同时训练阶段所需时间通常远超推理阶段。因此，现有太赫兹NOMA系统的资源分配方法仍存在以下核心不足：一是传统优化方法计算复杂度高，难以满足实时通信需求；二是基于有标签学习的方法依赖高质量最优标签，训练成本较高；三是现有强化学习方法在瞬时性能优化方面仍存在不足。

为解决现有太赫兹NOMA网络资源分配方法在计算复杂度高、最优标签依赖度高及瞬时性能优化能力不足的局限，本文提出一种新颖的基于Transformer结构的双Q网络模型，用于实现太赫兹非正交多址接入系统中的即时功率分配。该方法旨在在每个通信时隙内最大化系统总吞吐量，同时兼顾实时性与稳定性。双Q网络作为深度Q网络的重要改进结构，通过引入独立的目标Q网络有效缓解传统Q学习中Q值高估的问题，从而提升训练稳定性和策略收敛性[18][19]。为了进一步保障系统中的用户公平性，本文结合太赫兹通信场景中的长用户中心窗口原则[11]，即优先将中心带宽资源分配给远距离用户组，以弥补其信道条件劣势，而将边缘带宽分配给其他用户，从而实现性能与公平性的平衡。在此基础上，双Q学习算法按照既定用户顺序进行功率分配决策。经过仿真训练后，所提出模型能够在毫秒级时间尺度内快速生成稳定且自适应的次优功率分配方案，为太赫兹NOMA系统的即时资源管理提供了一种具有实际应用潜力的解决思路。

相比现有非机器学习的传统方法、有标签深度学习方法以及强化学习方法，本文所提出的太赫兹NOMA系统资源分配算法的主要贡献总结如下：

● 提出融合Transformer与双Q网络的资源分配框架：不同于传统强化学习方法，本文引入Transformer结构以增强对多用户信道状态特征的建模能力，从而更有效地刻画用户间复杂关系。同时，双Q网络的引入提高了训练的稳定性。

● 实现面向瞬时性能优化的即时功率分配策略：区别于现有强化学习方法侧重长期平均性能优化，本文以每个通信时隙的系统吞吐量最大化为目标，使模型能够在动态信道环境下实现快速响应与实时决策。

● 在性能与复杂度之间实现有效权衡：所提出模型在训练完成后能够在毫秒级时间内生成稳定的功率分配方案，并接近最优方案。相比基于非机器学习的传统方法显著降低计算复杂度，同时避免了有标签深度学习方法对采集高质量最优标签的依赖。

1　太赫兹非正交多址接入系统模型

本文考虑一个下行太赫兹非正交多址接入通信系统，其系统结构如图1所示。该系统由一个基站和若干用户设备构成，用户在基站覆盖范围内随机分布，不同用户与基站之间的距离、信道条件以及传播环境存在显著差异，从而形成具有代表性的太赫兹通信场景。基站与用户均采用MIMO技术，以提升传播距离与通信质量[20][21]。为兼顾系统性能与硬件实现复杂度，采用文献[22]中的混合波束赋形技术，即通过模拟波束赋形与数字预编码相结合的方式进行信号处理，从而在维持较低的制造成本和功耗的同时保持较高的波束成形增益。具体而言，基站配置线性天线阵列，通过混合预编码结构生成M个定向波束，并在不同波束中复用K个太赫兹子频段，实现空间与频谱资源的联合利用。考虑到太赫兹频段传播特性（即其路径损耗受到大气分子吸收效应影响），采用如下路径损耗模型：

$$\mathrm{PL}_{\mathrm{THz}}(d,f)=\left(\frac{4\pi fd}{c}\right)^2\exp\left(g(f)d\right)\#(1)$$

其中，c为光速，f是子载波频率，d为通信距离，$g(f)$则表示由大气分子吸收造成的频率相关衰减系数。经过发射端混合预编码矩阵$\boldsymbol{W}$和接收端混合处理矩阵的共轭转置矩阵$\boldsymbol{C}^*$处理后，每个波束对应的有效信道响应由原始信道矩阵$\boldsymbol{H}(\mathrm{PL}_{\mathrm{THz}}(d,f))$转变为：

$$h(d,f)=\boldsymbol{C}^*\boldsymbol{H}\left(\mathrm{PL}_{\mathrm{THz}}(d,f)\right)\boldsymbol{W}\#(2)$$

在该系统中，每个波束可同时为多个用户提供服务。受限于非正交多址接入接收端处理能力及系统复杂度要求[23]，每个波束内仅服务最多4

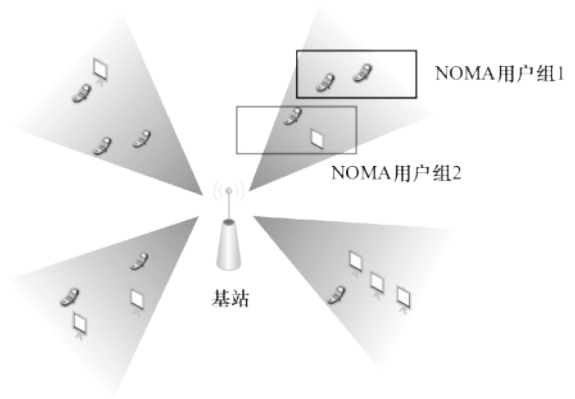

图1　太赫兹非正交多址接入系统

个用户。另一方面，太赫兹频段在某些特定频率附近存在明显的大气分子吸收峰，使得通信频谱呈现出若干可用“窗口”[24][25]。通常情况下，频谱中心子频段受吸收影响较小，信道条件较好，而边缘子频段则衰减更为严重。为了提升系统公平性并充分利用不同频段特性，本文依据用户与基站之间的距离对用户进行分组，将每个波束中的4个用户划分为两个NOMA用户组。其中，距离基站较远的两个用户分配至中心子频段，以补偿其路径损耗劣势；其余两个用户则分配至边缘子频段，从而在保证总体系统性能的同时兼顾远近用户的服务质量。

通过串行干扰消除技术，非正交多址接入系统能够按照用户信号功率大小顺序逐步解码信号。一般而言，接收端首先检测功率较大的用户信号，并在解码后将其从接收信号中消除，再对剩余功率较小的用户信号进行解码[26][27]。因此，对于第m波束中的第i个NOMA用户组，若两位用户在第k子频段所分配的功率分别是P_1和P_2，且$P_1<P_2$，则在此子频段的信干噪比（signal-to-interference-plus-noise ratio，SINR）分别为：

$$\gamma_1(m,k,i)=\frac{P_1(m,k,i)\left|\widetilde{h_1}(m,k,i)\right|^2}{\widetilde{I_1}(m,k,i)+\sigma^2}\#(3)$$

$$\gamma_2(m,k,i)=\frac{P_2(m,k,i)\left|\widetilde{h_2}(m,k,i)\right|^2}{\widetilde{I_2}(m,k,i)+P_1(m,k,i)\left|\widetilde{h_2}(m,k,i)\right|^2+\sigma^2}\#(4)$$

其中，$\tilde{I}$为去除自干扰之后的干扰强度，σ^2为混合波束赋形处理后的等效噪声功率。$\widetilde{h_1}$和$\widetilde{h_2}$分别代表P_1和P_2所对应用户的有效信道响应，其值根据子载波频率以及用户与基站距离可由式（2）得出。

因此，对于所考虑的太赫兹非正交多址接入网络，在满足每个用户最小速率$R_{\min}$和基站最大发射功率约束$P_{\max}$的前提下，系统功率分配方案$\boldsymbol{P}$的优化目标可以表示为最大化系统总吞吐量，即：

$$\max_{\boldsymbol{P}}\ \sum_{m=1}^{M}\sum_{k=1}^{K}\sum_{i=1}^{2}R_1(m,k,i)+R_2(m,k,i)\#(5)$$

$$\text{s.t.}\ \sum_{k=1}^{K}R_1(m,k,i)\geq R_{\min},\sum_{k=1}^{K}R_2(m,k,i)\geq R_{\min},\forall m,i,\#(6)$$

$$\sum_{m=1}^{M}\sum_{k=1}^{K}\sum_{i=1}^{2}P_1(m,k,i)+P_2(m,k,i)\leq P_{\max}\#(7)$$

2　基于Transformer的双Q网络功率分配方法

针对上节提出的高维非凸功率分配问题，传统优化方法计算复杂度高，难以满足实时性需求。为此，本文引入深度强化学习，将多用户功率分配建模为序贯决策过程，使系统能够在动态信道环境下逐步生成近似最优的策略。考虑到不同用户与子频段间的资源分配存在耦合关系，本文采用Transformer网络提取功率分配中的全局关联特征。得益于自注意力机制对长距离依赖的建模能力，Transformer在复杂序贯决策任务中表现出优越性能[28][29]。

在本文强化学习框架中，基站作为智能体，太赫兹NOMA网络作为环境。智能体根据当前状态选择功率分配动作，环境反馈即时奖励并更新

状态，该过程可建模为马尔可夫决策过程。通过持续交互，智能体能够在满足功率约束与用户速率需求的前提下逐步提升系统吞吐量。然而，在大规模状态与动作空间下，传统非深度强化学习方法易受到维数灾难影响。为此，本文采用深度Q网络处理连续状态，并进一步引入双Q网络结构，即通过增加独立目标网络提升训练稳定性[18][30]。

2.1　基于 Transformer 的双 Q 网络结构与训练流程

为了使模型能够充分关注当前用户分布和信道条件，并学习不同用户之间功率分配的关联特性，本文提出的深度强化学习方法采用分步决策机制，即在一次训练过程中按既定顺序逐步输出全网功率分配决策。具体而言，每一步决策仅为某一NOMA用户组在特定子频段上的功率分配生成结果，并依据已完成的分配状态决定下一步需要处理的用户组或子频段，直至完成所有用户组在全部子频段上的功率分配。这种序贯决策方式既降低了单次决策维度，又有利于Transformer捕捉不同分配步骤之间的依赖关系。

如图2所示，本文所采用的双Q网络由在线Q网络和目标Q网络两部分组成，两者具有相同的Transformer网络结构。在线Q网络用于根据当前环境状态输出不同功率分配动作对应的Q值（即未来累计收益的期望），并选择Q值最大的动作作为当前最优决策。在训练过程中，在线Q网络通过与环境交互获得状态转移样本，并利用反向传播算法不断更新网络参数。Q网络的更新目标符合贝尔曼方程，即当前时刻的Q值Q_t等于当前环境状态s_t下最优动作a_t的收益r_t加上下一个时刻（即$t+1$时刻）的Q值。然而，由于在线网络同时负责动作选择和价值评估，容易产生Q值过高估计问题[30]。本文引入目标Q网络用于计算目标Q值，其参数周期性从在线网络复制更新（即在线Q网络更新数次之后，目标Q网络更新一次），从而降低估计偏差并增强训练稳定性。为了实现上述双Q网络的功能，Transformer网络结构的输入为环境状态，输出为不同潜在资源分配动作所对应的Q值。本文采用经典的Transformer结构，仅针对太赫兹NOMA网络的环境状态和资源分配动作适应性改造输入层和输出层维度。对应的损失函数可表示为。神经网络的损失函数可以设计为：

$$\mathrm{Loss}(\theta)=E\left[\left(r_t+\gamma' Q\left(s_{t+1},\mathrm{argmax}_{a_{t+1}'}Q\left(s_{t+1},a_{t+1}';\theta\right);\hat{\theta}\right)-Q\left(s_t,a_t;\theta\right)\right)^2\right]\#(8)$$

其中，θ为在线Q网络的参数，$\hat{\theta}$是目标Q网络的参数，$\gamma'\in[0,1]$为未来奖励折扣因子，表示对未来动作奖励值的重视情况（0为完全看重现有动作，1为现有动作和未来动作一样重要。本文取1，以强调总收益与即时收益同等重要，从而更好地优化系统整体吞吐量）。

在训练阶段，智能体采用ε-贪婪策略在探索与利用之间取得平衡。具体而言，智能体以概率ε随机选择功率分配动作以探索未知策略空间，而以概率1-ε根据当前Q网络输出的最优动作进行决策。随着训练进行逐渐降低ε值，使模型在早期充分探索以避免陷入局部最优，在后期则更加侧重利用已学习到的高价值策略[31]。此外，引入经验回放机制，将智能体与环境交互产生的状态转移样本存储至经验池，并在训练时随机抽取小批量样本更新网络参数，从而降低样本相关性，提高训练稳定性与收敛速度[32]。

经过离线训练后，该基于Transformer的双Q网络模型可直接应用于不同用户分布及信道条件的在线功率分配场景，无需重新训练即可快速输出分配策略，体现出较强的泛化能力和良好的实时决策性能。

2.2 马尔可夫决策过程建模

为将功率分配问题映射到上述强化学习框架中，需要对图2中马尔可夫决策过程的状态空间、动作空间以及奖励函数进行合理设计，这些要素共同决定了强化学习算法的收敛性能及最终策略质量。状态空间应能够全面反映当前无线环境特征及资源分配情况，动作空间对应可选的功率分配决策，而奖励函数用于引导智能体朝着性能优化方向学习。

● 状态空间：本文将系统状态定义为由等效信道增益信息、当前待分配功率的用户组与子频段序号以及已有功率分配状态共同构成的状态向量。该设计能够全面反映不同用户信道条件差异、频谱资源使用情况以及当前系统分配进度，使智能体能够基于充分的环境信息做出合理决策。

● 动作空间：考虑到连续功率分配会显著增加强化学习算法的学习难度与计算复杂度，本文对可分配功率进行离散化处理，将其划分为L个等差功率等级$\{p_1, p_2, \ldots, p_L\}$，$p_1=0<p_2<\ldots<p_L$。其中，中间功率等级设定为平均分配对应的值，即$p_{\lfloor L+1 \rfloor}=\frac{P_{\max}}{2MK}$。在每一步决策中，智能体为指定NOMA用户组及其所使用的某一子频段选择一个功率等级，逐步完成全系统功率分配。这种离散化策略在保证决策灵活性的同时有效降低了动作空间维度，使强化学习算法更易收敛。

● 奖励函数：奖励函数的设计直接影响强化学习算法的收敛速度和最终性能。本文以系统总吞吐量的增量作为即时奖励，鼓励智能体优先选择能够提升整体系统吞吐量的功率分配方案。由于吞吐量数值通常较大，为保证训练稳定性，本文对奖励进行归一化处理，即将吞吐量除以1×10^{10}。同时，为确保功率分配结果满足系统约束条件，在奖励函数中引入惩罚机制：当系统总发射功率超过最大功率限制时，给予智能体−10的负奖励，并按比例缩小所有功率分配值直至满足功率约束，以避免产生不可行的解；当功率分配完成后若存在用户速率低于最小速率要求，则施加更严格惩罚，即给予−100奖励值，以强化模型

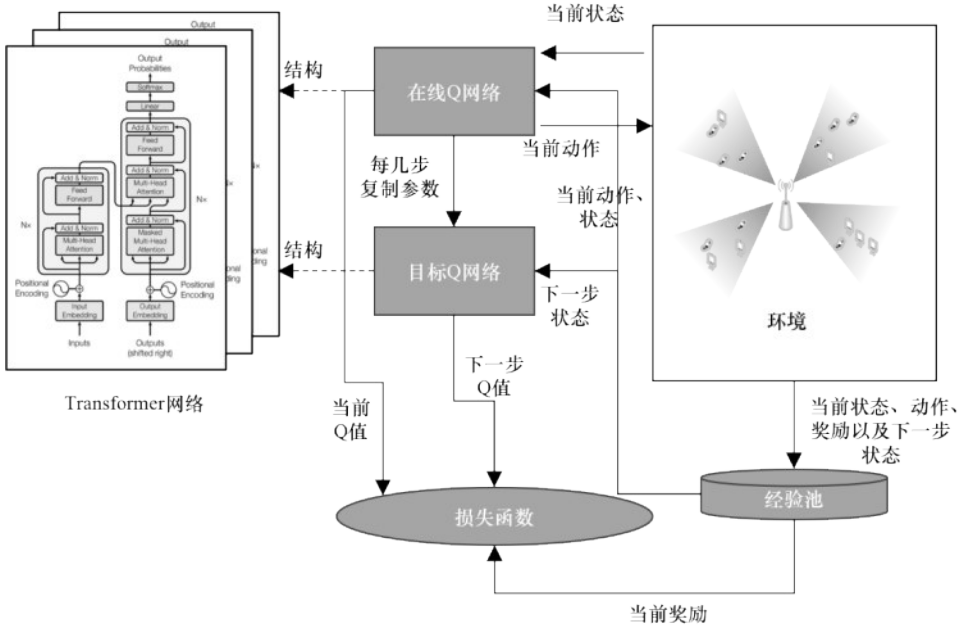

图2 基于Transformer的双Q网络与其决策过程

对服务质量约束的重视。通过奖励与惩罚机制的协同设计，可有效引导智能体在提升吞吐量的同时兼顾功率限制与用户公平性需求。

3　数值结果与分析

本节通过仿真实验验证所提出基于 Transformer 的双 Q 网络功率分配算法在太赫兹 NOMA 系统中的有效性与工程可行性。仿真实验中，16 个用户设备随机分布在基站周围半径 30 m 的覆盖区域内。用户端均采用由 128 个 8×1 天线子阵列构成的混合波束赋形结构，基站端则配置由 256 个 8×1 子阵列组成的更大规模阵列。系统采用三个可用子频段进行通信，频率范围分别为 277.5 – 292.5 GHz、292.5 – 307.5 GHz 和 307.5 – 322.5 GHz，覆盖了太赫兹通信中典型的可用传输窗口。为保证公平性和服务质量，基站最大发射功率约束设定为 20 dBm，同时要求每个用户的最低传输速率不低于 10 Gbps。本文仿真实验在配备 Intel i9-14900KF CPU、NVIDIA RTX 2080Ti GPU 及 32GB 内存的计算平台上完成，模型基于 PyTorch 框架实现。3.1 神经网络超参数设计

在本文提出的基于 Transformer 的双 Q 网络模型中，训练阶段采用 ε-贪婪策略控制探索与利用的权衡。初始探索率设定为 ε=1，即训练早期完全探索，以避免策略陷入局部最优；随后随着训练推进，ε 按 0.997 的衰减率逐步降低，直至小于 0.001，使模型在训练后期更多依赖已学习到的高价值动作，提高收敛效率与策略稳定性。整个训练过程包含 10 000 轮，且训练数据集由 10 000 个随机样本组成。每轮训练会随机抽取一个样本更新模型参数。系统采用容量为 2 000 的经验回放池，采用先入先出（first in, first out，FIFO）机制：即当经验池满时，新样本将自动替换最早存入的旧样本。

在深度强化学习训练中，学习率与批量大小是影响收敛速度与最终性能的重要超参数。学习率决定了梯度下降中参数更新的幅度。过高可能导致模型越过最优解，过低则收敛缓慢。批量大小则决定了每次梯度更新所使用的样本数量，是深度学习中的关键超参数。本文对比了不同学习率和批量大小的训练效果。如图 3 所示，0.001 的学习率和 16 的批量大小收敛速度更快，且收敛之后结果更稳定。综合考虑训练效率与稳定性，本文后续实验均采用学习率 0.001、批量大小 16 的超参数配置。

3.2　算法性能

图 4 展示了不同算法在训练过程中系统吞吐量随训练轮次变化的曲线，并将穷举搜索算法的最优总速率作为性能上界进行对比。可以观察到，随着训练周期增加，本文所提出的算法吞吐量虽存在一定波动，但整体呈上升趋势，并逐渐逼近穷举搜索的最优结果，最终收敛到稳定水平。更重要的是，在相同训练轮次下，本文所提出算法的吞吐量显著优于基于 Transformer 的深度 Q 网络以及无 Transformer 的传统双 Q 网络和深度 Q 网络（即基于多层感知器的传统双 Q 网络和深度 Q 网络）。由图可知，双 Q 机制增强了训练稳定性，使得模型在学习过程中更易取得显著的性能提升。此外，Transformer 结构对多用户、多子频段之间关联关系的刻画能够有效提升策略学习效率与决策质量，使得算法的性能进一步上升。

为了进一步评估训练完成后的泛化能力与在线推理效率，本文将训练完成的模型用于 100 个随机生成的测试用例。在所有测试用例中，本文算法均未出现用户速率不达标现象，表明所设计的奖励惩罚机制与约束处理方式能够有效引导策略生成满足 QoS 需求的可行解。同时，测试用例的平均吞吐量、平均频谱效率以及平均能量效率如表 1 所示。实验表明，本文提出的基于 Transformer 的双 Q 网络模型平均吞吐量、频谱效率以及能量效率可达到最优水平的 98% 以上，远高于传统双 Q 网络、基于 Transformer 的深度 Q 网络及

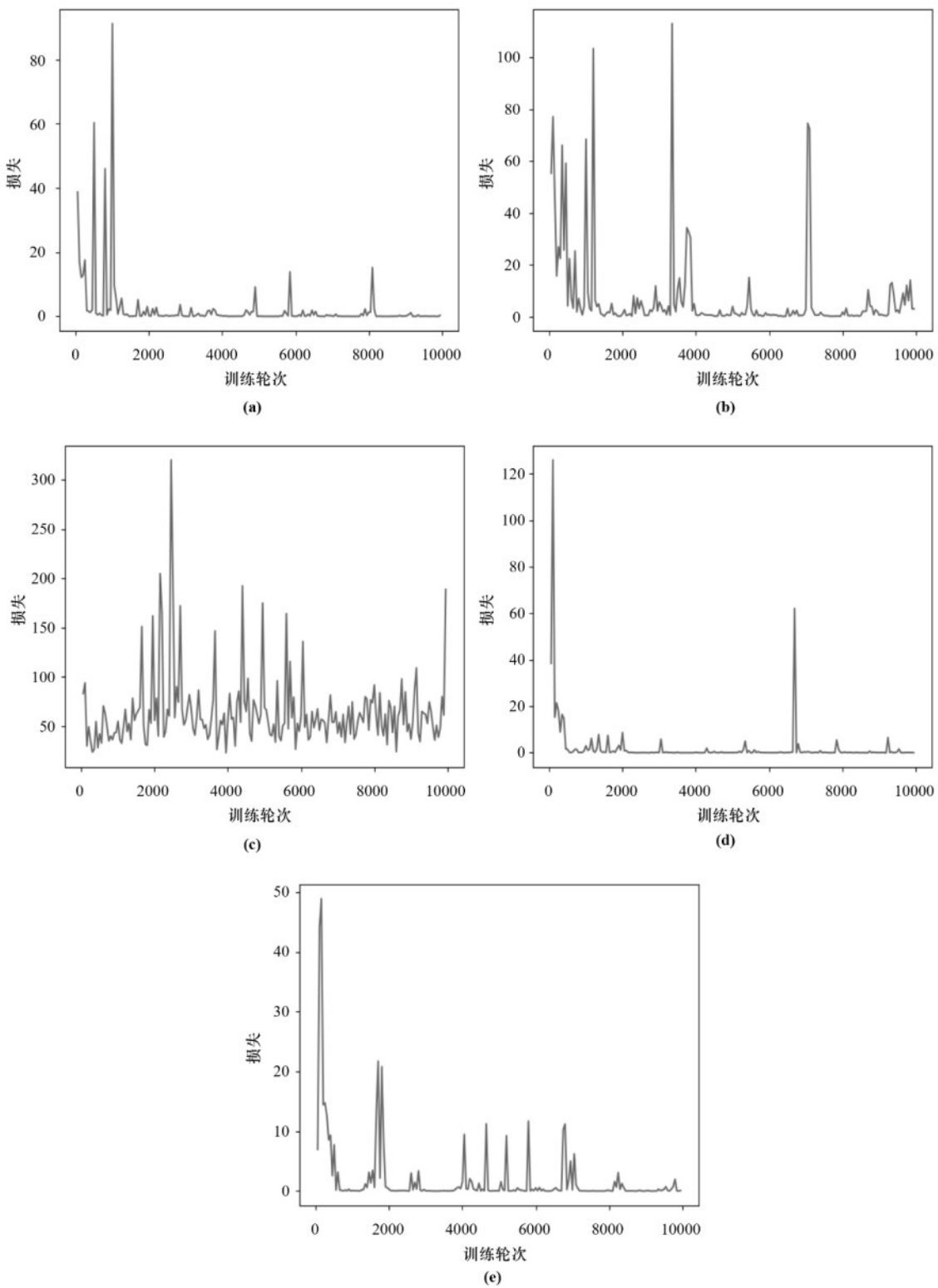

图3　不同学习率和批量大小下的训练过程：(a) 学习率：0.001，批量大小：16；(b) 学习率：0.01，批量大小：16；(c) 学习率：0.1，批量大小：16；(d) 学习率：0.001，批量大小：32；(e) 学习率：0.001，批量大小：8

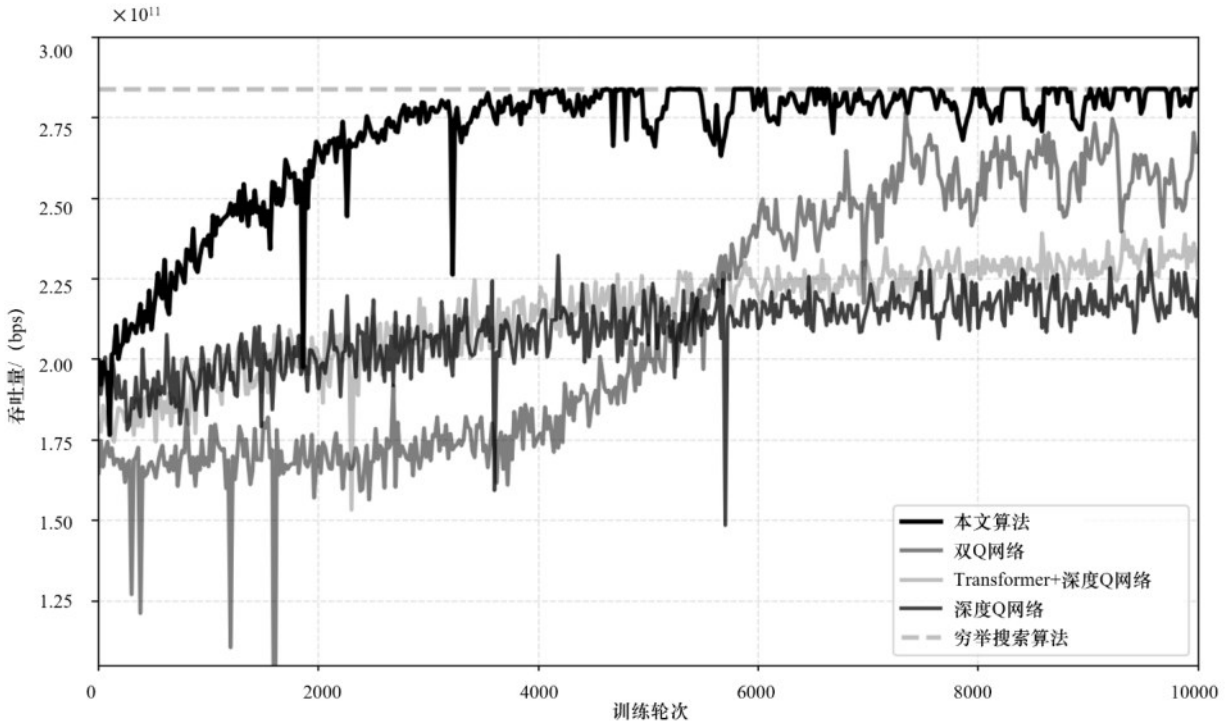

图4　算法吞吐量性能对比

传统深度Q网络算法。

同时，本文采用离线训练策略，即将离线训练好的模型应用于在线场景，不再在线训练。因此，相较于训练时长，本文更加关注训练好的模型应用时的推理时间。如表1所示，本文算法运行时间仅为最优方法的1.3%，平均每次输出仅需6 ms（单个样本的平均推理耗时的统计方式为：重复运行100个随机测试样例，每次输入单个样本，并将总耗时除以样本数量）。

此外，在深度强化学习算法模型隐藏层神经元数量和迭代次数不变的情况下，模型计算复杂度仅与输入和输出维度成正比例（即与波束数量M，每个波束内可复用的子频段数量K，分配功率的等级数量L成正比例）。因此，深度强化学习算法的训练和推理复杂度为$O(MKL)$。而穷举搜索法需对每个用户和子频段的组合尝试分配L个功率等级。本文中每个子频段在每个波束内均服务两位用户。总共有$2MK$个用户和子频段的组合。因此，穷举搜索法的复杂度为$O\left(L^{2MK}\right)$。因此，包括本文中在内的深度强化学习算法复杂度随用户规模线性增长，明显优于穷举搜索法的指数级增长，实际部署可行性明显更高。

本文进一步比较了用户公平性（采用无线网络中常用的Jain’s Fairness Index指标[23]：取值在0~1之间，并且越接近1，公平性越高）。如表1所示，本文所提出算法在取得接近穷举搜索的最优吞吐量的同时，公平性显著提升，并与其他对比算法的公平性表现相当，体现了良好的吞吐量与公平性的性能平衡。

进一步对测试数据进行统计分析发现，该算法下不同测试用例的结果和最优结果的比值波动幅度非常小。如图5所示，比值的分布曲线几乎与正态分布重合（均值为0.983，标准差为0.00226）。这种小标准差的高斯分布特征表明，在不同用户分布、不同信道条件下，本文算法不仅能够长期保持接近最优的吞吐量水平，而且性能稳

表1　性能对比（吞吐量、频谱效率、能量效率、用时、公平性）

性能	本文算法	双Q网络	Transformer +深度Q网络	深度Q网络	穷举搜索
相对最优值的平均吞吐量百分比	98.3%	91.5%	80.8%	77.5%	100%
平均频谱效率	6.2 bit/s/Hz	5.8 bit/s/Hz	5.1 bit/s/Hz	4.9 bit/s/Hz	6.3 bit/s/Hz
平均能量效率	2.79×10^{12} bit/J	2.68×10^{12} bit/J	2.34×10^{12} bit/J	2.26×10^{12} bit/J	2.84×10^{12} bit/J
测试用例总耗时	0.60 s	0.46 s	0.52s	0.33 s	44.96 s
公平性	0.62	0.61	0.65	0.66	0.51

定、可靠性强，具有较好的环境适应性与工程部署价值。

4　结束语

本文针对太赫兹非正交多址接入系统中功率分配问题计算复杂度高、难以满足动态场景实时调度需求的挑战，提出了一种基于Transformer和双Q网络的即时功率分配方法。在太赫兹通信环境中，由于频段高、信道变化敏感以及用户数量不断增加，传统依赖迭代优化或穷举搜索的功率分配算法往往计算量巨大、响应速度较慢，难以适应实时无线通信系统的调度需求。为此，本文将功率分配过程建模为马尔可夫决策过程，利用强化学习框架对动态资源分配问题进行建模与求解，并通过融合基于Transformer的双Q网络结构逐步为不同用户分配功率，从而在满足系统功率约束和用户服务质量约束条件下实现高效、快速且稳定的资源分配决策。

仿真结果表明，所提出的方法在系统总吞吐量方面能够逼近穷举搜索所得的最优解，同时显著降低了算法计算复杂度。相比传统优化方法，该模型在多用户太赫兹非正交多址接入场景中能够实现毫秒级功率分配决策，具备较高的实时性，能够满足未来高速通信系统对快速调度与低时延响应的要求。此外，该方法无需在每次决策时重新执行复杂优化过程，一旦离线训练完成，

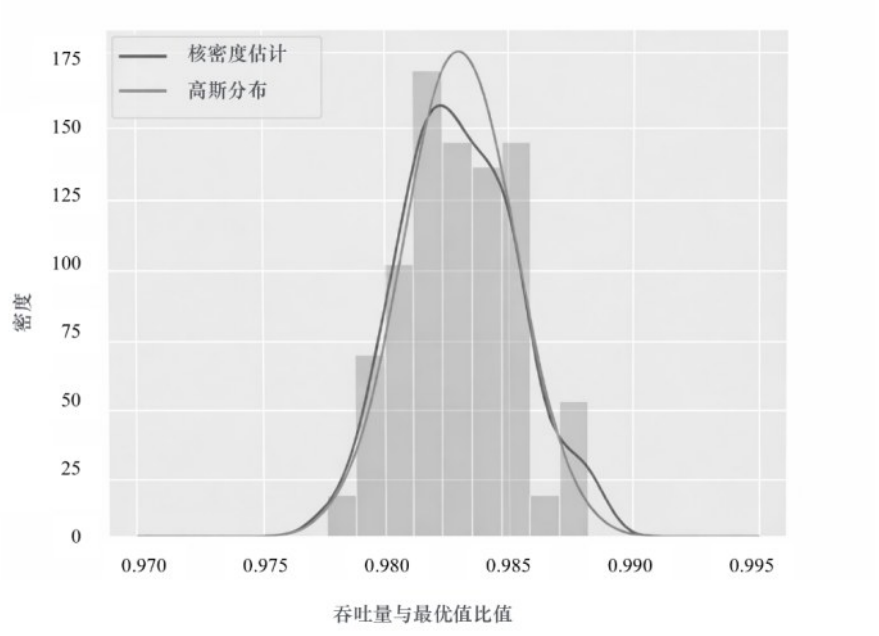

图5　吞吐量分布

模型即可直接输出功率分配策略，从而有效减少在线计算负担，具备较强的工程应用潜力。此外，本文提出的双Q网络框架不依赖于预先计算得到的最优功率分配标签，避免了传统监督学习方法中标签获取困难、计算代价高的问题。通过强化学习机制，模型能够在交互过程中自主学习功率分配策略，并在不同用户分布、信道衰落条件以及系统参数变化情况下保持较好的泛化能力和环境适应性。未来工作可进一步考虑更复杂的系统模型，例如联合波束赋形、子带分配与功率控制的端到端学习方案，以及在用户高速移动和信道快速变化条件下的鲁棒性问题，以进一步提升太赫兹非正交多址接入系统的整体性能。

参考文献：

[1] 唐雄燕,李福昌,张忠皓,等. 6G网络需求,架构及技术趋势[J]. 移动通信, 2021, 45(4):37-44.
TANG X Y, LI F C, ZHANG Z H, et al. Requirements, architectures and technology trends of 6G network[J]. Mobile Communications, 2021, 45(4): 37–44.

[2] WANG C X, YOU X, GAO X, et al. On the road to 6G: visions, requirements, key technologies, and testbeds[J]. IEEE Communications Surveys & Tutorials, 2023, 25(2): 905-974.

[3] 王战将,李凯乐,张飞翔,等. 面向6G的太赫兹通信技术研究[J]. 移动通信, 2025, 49(5):121-127.
WANG Z J, LI K L, ZHANG F X, et al. A review of terahertz communication technology for 6G[J]. Mobile Communications, 2025, 49(5): 121–127.

[4] PETROV V, PYATTAEV A, MOLTCHANOV D, et al. Terahertz band communications: applications, research challenges, and standardization activities[C]//Proceedings of 2016 8th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT). Piscataway: IEEE Press, 2016: 183-190.

[5] AKYILDIZ I F, HAN C, HU Z, et al. Terahertz band communication: an old problem revisited and research directions for the next decade[J]. IEEE Transactions on Communications, 2022, 70(6): 4250-4285.

[6] XUE Q, JI C, MA S, et al. A survey of beam management for mmWave and THz communications towards 6G[J]. IEEE Communications Surveys & Tutorials, 2024, 26(3): 1520-1559.

[7] DAI L, WANG B, YUAN Y, et al. Non-orthogonal multiple access for 5G: solutions, challenges, opportunities, and future research trends[J]. IEEE Communications Magazine, 2015, 53 (9): 74-81.

[8] SAITO Y, KISHIYAMA Y, BENJEBBOUR A, et al. Non-orthogonal multiple access (NOMA) for cellular future radio access[C]//Proceedings of 2013 IEEE 77th Vehicular Technology Conference (VTC Spring). Piscataway: IEEE Press, 2013: 1-5.

[9] AHMED A, WANG X, HAWBANI A, et al. Unveiling the potential of NOMA: a journey to next-generation multiple access [J]. IEEE Communications Surveys & Tutorials, 2024, 27(5): 3099-3164.

[10] DING Z, ADACHI F, POOR H V. The application of MIMO to non-orthogonal multiple access[J]. IEEE Transactions on Wireless Communications, 2016, 15(1): 537-552.

[11] ZHANG X, HAN C, WANG X. Joint beamforming-power-bandwidth allocation in terahertz NOMA networks[C]//Proceedings of 2019 16th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON). Piscataway: IEEE Press, 2019: 1-9.

[12] JIAO L, SHAO Y, SUN L, et al. Advanced deep learning models for 6G: overview, opportunities, and challenges[J]. IEEE Access, 2024, 12: 133245-133314.

[13] MAHMOOD M R, MATIN M A, SARIGIANNIDIS P, et al. A comprehensive review on artificial intelligence/machine learning algorithms for empowering the future IoT toward 6G era[J]. IEEE Access, 2022, 10: 87535-87562.

[14] SANJALAWE Y, FRAIHAT S, ABUALHAJ M, et al. A review of 6G and AI convergence: enhancing communication networks with artificial intelligence[J]. IEEE Open Journal of the Communications Society, 2025, 6: 2308-2355.

[15] SAETAN W, THIPCHAKSURAT S. Power allocation for sum rate maximization in 5G NOMA system with imperfect SIC: a deep learning approach[C]//Proceedings of 2019 4th International Conference on Information Technology (InCIT). Piscataway: IEEE Press, 2019: 195-198.

[16] XIAO L, LI Y, DAI C, et al. Reinforcement learning-based NOMA power allocation in the presence of smart jamming[J]. IEEE Transactions on Vehicular Technology, 2018, 67(4): 3377-3389.

[17] ZHANG Y, WANG X, XU Y. Energy-efficient resource allocation in uplink NOMA systems with deep reinforcement learning [C]//Proceedings of 2019 11th International Conference on Wireless Communications and Signal Processing (WCSP). Piscataway: IEEE Press, 2019: 1-6.

[18] VAN H H, GUEZ A, SILVER D. Deep reinforcement learning

with double Q-learning[C]//Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence (AAAI-16). Washington D.C: AAAI Press, 2016: 2094–2100.

[19] ZHAO D, LIU D, LEWIS F L, et al. Special issue on deep reinforcement learning and adaptive dynamic programming[J]. IEEE Transactions on Neural Networks and Learning Systems, 2018, 29(6): 2038-2041.

[20] 李伟琨,姚信威,王万良,等. 太赫兹通信中MIMO信道建模与容量分析[J]. 计算机工程, 2015, 41(4):4.
LI W K, YAO X W, WANG W L, et al. MIMO channel modeling and capacity analysis in terahertz communication[J]. Computer Engineering, 2015, 41(4): 4.

[21] ALHAJ N A, JAMLOS M F, MANAP S A, et al. Integration of hybrid networks, AI, ultra massive-MIMO, THz frequency, and FBMC modulation toward 6G requirements: a review[J]. IEEE Access, 2023, 12: 483-513.

[22] YAN L, HAN C, YUAN J. Energy-efficient dynamic-subarray with fixed true-time-delay design for terahertz wideband hybrid beamforming[J]. IEEE Journal on Selected Areas in Communications, 2022, 40(10): 2840-2854.

[23] HU Z, HAN C, DENG Y, et al. Multi-task deep reinforcement learning for terahertz NOMA resource allocation with hybrid discrete and continuous actions[J]. IEEE Transactions on Vehicular Technology, 2024, 73(8): 11647-11663.

[24] HAN C, WANG Y, LI Y, et al. Terahertz wireless channels: a holistic survey on measurement, modeling, and analysis[J]. IEEE Communications Surveys & Tutorials, 2022, 24(3): 1670-1707.

[25] WANG J, WANG C X, HUANG J, et al. 6G THz propagation channel characteristics and modeling: recent developments and future challenges[J]. IEEE Communications Magazine, 2022, 62(2): 56-62.

[26] ISLAM S M R, AVAZOV N, DOBRE O A, et al. Power-domain non-orthogonal multiple access (NOMA) in 5G systems: potentials and challenges[J]. IEEE Communications Surveys & Tutorials, 2016, 19(2): 721-742.

[27] HE C, HU Y, CHEN Y, et al. Joint power allocation and channel assignment for NOMA with deep reinforcement learning[J]. IEEE Journal on Selected Areas in Communications, 2019, 37 (10): 2200-2210.

[28] WEN M, LIN R, WANG H, et al. Large sequence models for sequential decision-making: a survey[J]. Frontiers of Computer Science, 2023, 17(6): 176349.

[29] YUAN W, CHEN J, CHEN S, et al. Transformer in reinforcement learning for decision-making: a survey[J]. Frontiers of Information Technology & Electronic Engineering, 2024, 25(6): 763-790.

[30] 陈卓, 冯钢, 何颖, 等. 运营商网络中基于深度强化学习的服务功能链迁移机制[J]. 电子与信息学报, 2020, 42(9): 2173-2179.
CHEN Z, FENG G, HE Y, et al. Deep reinforcement learning based migration mechanism for service function chain in operator networks[J]. Journal of Electronics & Information Technology, 2020, 42(9): 2173–2179.

[31] TERMEHCHI A, BAO T, SYED A, et al. Goal-oriented reinforcement learning in THz-enabled UAV-aided network using supervised learning[J]. IEEE Open Journal of the Communications Society, 2024, 5: 5027-5036.

[32] KHALILI A, MONFARED E M, ZARGARI S, et al. Resource management for transmit power minimization in UAV-assisted RIS HetNets supported by dual connectivity[J]. IEEE Transactions on Wireless Communications, 2021, 21(3): 1806-1822.